Fluctuations and large deviationsin game theoretical models

Nicole, Robin Maxime Jacques

Awarding institution:
King's College London

KING'S College LONDON

Doctoral Thesis

Fluctuations and large deviations in game theoretical models

Author

Robin Nicole

Supervisor

Prof. Peter Sollich

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in the

Disordered Systems group, Department of Mathematics

December 11, 2017

Abstract

This thesis uses advanced methods from statistical physics to investigate collective effects in game theoretical models with a large number of players. Specifically, I study the strategy distributions in a large game that models how agents choose among double auction markets. I provide a classification of the possible mean field Nash equilibria, which include potentially segregated states where an agent population can split into subpopulations adopting different strategies. I compare this classification with the results of Experience-Weighted Attraction (EWA) learning, which in the long run leads to Nash equilibria in the appropriate limits of large intensity of choice, low noise (long agent memory) and perfect imputation of missing scores (fictitious play). Non-trivially, depending on how the relevant limits are taken, more than one type of equilibrium can be selected. These include the standard homogeneous mixed and heterogeneous pure states, but also heterogeneous mixed states in which different agents play different strategies that are not all pure. I also investigate the influence of heterogeneity in traders' behaviour on the emergence of segregation described previously. The theoretical machinery is then extended from the study of systems with two markets, to the analysis of multi-agent systems where traders can choose between multiple markets.

The last part of the thesis focuses on how the interaction among players in a repeated game can be represented as an evolutionary process in a "population of ideas". I propose the interpretation of reinforcement learning as a stochastic

process in a finite population of this type. The resulting birth-death dynamics has absorbing states and allows for the extinction of ideas, which marks a key difference with mutation-selection processes. I characterise the outcome of evolution in populations of ideas for several classes of symmetric and asymmetric games.

Acknowledgements

First, I would like to thank Peter Sollich for his enlightening advice as well as his support during those four years. Thanks, as well to Tobias Galla for his supervision and his willingness to share his passion for science during my short but enriching stay in Manchester.

This thesis would not have been possible without the support from all the Disordered systems group, thanks to all of you for sharing your knowledge during the Monday seminar. Thanks as well to the KCL Department of mathematics for funding this Ph.D. and for providing me such a pleasant working environment.

I would also like to thank all the fellow graduate students: Anshul, Andrea, Carla, Celeste, Davide, Dimitris, Michaella, Michele, Nicolas, Pablo, Pierre, Riccardo, Ryan, Sam, and Victor for all the good time we have spent together.

Last but not least I would like to thank my family, as well as my friends from the other side of the Channel for their constant support during this four-year journey.

Contents

List of Figures

List of Tables

Previously published work

(i) Robin Nicole and Peter Sollich
Dynamical selection of Nash equilibria using Experience Weighted Attraction Learning: emergence of heterogeneous mixed equilibria
Under review

(ii) Robin Nicole, Peter Sollich, and Tobias Galla
Stochastic evolution in populations of ideas.
Scientific Reports, 7:40580, 2016.

Paper (i) covers the content of Chap. 2 and paper (ii) covers the content of Chap. 5.

CHAPTER 1

Introduction

1.1 A really brief introduction to game theory

Game theory aims at the "study of mathematical models of conflicts and cooperation between intelligent, rational decision-makers"[105]. Given such a broad definition, it is not surprising that game theory has many different fields of application. These range from the design of efficient market mechanisms [64] to biology where it is used to model population dynamics [61, 50, 34] and even the development of efficient policies [62] and the design of multi-agent systems [107].

To give a brief introduction to *Game theory* and to the concept of Nash equilibrium, let us consider the well-known *prisoners' dilemma* where two bandits – called *players* in game theory – are arrested and suspected of having committed a crime. They are placed in two distinct interrogation rooms, cannot communicate and are faced with two choices: betraying their fellow or remaining silent. If they both remain silent, they will both be imprisoned for one year, and if they betray each other, they will both go to prison for five years. However, if one of them betrays the other while his fellow remains silent, the

prisoner who has betrayed will not be convicted, and the silent prisoner will be sentenced to ten years in prison. We can summarise this situation by the *payoff matrix* shown in Tab. 1.1. One question that arises is what strategies players are likely to play.

One possible approach to answering this question is to look for strategies corresponding to an "equilibrium". We could, for example, consider the *Nash equilibrium* [68] which generalises the notion of equilibrium – that is common in physics – to games such as the prisoners' dilemma. Its definition is: *a strategy profile (i.e. the list of probabilities for picking each of the available actions for each player) is said to be a Nash equilibrium if none of the players (here the bandits) can increase his payoff (here minus the number of years to be spent in jail) by changing strategy* unilaterally. One can also see the Nash equilibrium as the long run outcome of a game repeatedly played when each player plays the best response to the former strategy profile (this is the *best response dynamics*). For example in our game, as shown in the Tab. 1.1, a silent bandit will always decrease his jail sentence if he betrays. As a consequence, the only possible Nash equilibrium for the prisoners' dilemma is when both bandits betray. The notion of a Nash equilibrium is one possibility among many to define the equilibrium of *rational* players in games. (Un)fortunately, in practice, humans are not rational. They might be subject to biases [54, 103, 92], might not be capable of calculating their best strategy and end up not converging to their Nash equilibrium [70]. Thankfully, modeling *bounded rationality* is an active area of research of which we want to give a brief taste in the next paragraph.

One possible way to take into account bounded rationality is to consider players who do not follow best-response dynamics. For an extensive review of learning dynamics in games, we refer to [40]. Among the existing learning dynamics in games, *belief-based learning* where traders learn their beliefs about the probable actions taken by other players and *reinforcement-based learning*

Bandit 1 \ Bandit 2	Remain silent	Betray
Remain silent	$(-1, -1)$	$(-10, 0)$
Betray	$(0, -10)$	$(-5, -5)$

Table 1.1: Payoff matrix of the prisoners' dilemma game where each line corresponds to one strategy available for the first bandit and each column to one strategy available to the second bandit. The entries of the matrix are of the form (payoff of bandit 1, payoff of bandit 2) where the payoff of a bandit is minus the number of years he will spend in jail. The arrows (red for bandit one and black for bandit 2) show how a bandit would change his strategy under the best response dynamics. We see that the Nash equilibrium (Betray, Betray) is the only strategy profile that is stable under the best response dynamics.

where players reinforce their strategies based on the payoff they provide are some of the most interesting. In this thesis, we focus specifically in *Experience-Weighted Attraction*(EWA) learning, which is an interpolation between belief based learning and reinforcement based learning [19].

Another direction to go beyond the full rationality assumption is to consider evolutionary games, which model Darwinian selection. In evolutionary games, one no longer studies players who learn to play against each other, but instead populations of individuals that reproduce proportionally to how well-suited they are to the environment. The approach is rooted in the seminal paper of Maynard-Smith *et al.* [61] "The logic of animal conflict". We refer the interested reader to an extensive review of evolutionary game theory in the textbook of Hofbauer and Sigmund [50].

1.2 Mission statement

In brief, the aim of this thesis is to study the long run outcome of different learning dynamics in games using tools coming from statistical physics as well as large deviation methods. Such an approach makes it possible to obtain the long run outcome of EWA learning dynamic in limiting cases (players with large memory) and to calculate finite size quantities such as first passage times. More specifically we investigate rare events in learning dynamics and their impact on the shape of the distribution of strategies among players. Our particular motivation is to study *segregation*, a phenomenon whereby players spontaneously split into groups adopting different strategies. We pursue this question in the context of players choosing between multiple double auction markets, a setup that we model as an *aggregative game* [29] with a *large number of players*. Aggregative game here designates a game where the payoff of players only depends on their own strategy and some mean field quantities. These summarise the behaviour of the other players and are called aggregates. For example, in this thesis where we study segregation among several markets, the aggregates are the number of buyers and the number of sellers in each market. Another famous example of an aggregative game is the Cournot model of oligopoly [30]. This model is concerned with several firms who compete on the amount of a good they produce. In this model, the payoff of firm i is $q_i P(\bar{q}) - C(q_i)$ where $C(q_i)$ is the cost of production of q_i units of good and $P(\bar{q})$ is the price of a single unit of good. The price of a single unit of good depends on the total quantity of a good produced $\bar{q} = \sum_i q_i$ which is the aggregate we mentioned earlier. Since the payoff of the agents depends only on the aggregates of the agents, there can be many distributions of strategies among the players of an aggregative game that are Nash-equilibria. There is a vast literature on the dynamics of aggregates when players learn to play an aggregative game [55, 39]. However, to assess the existence of groups, one

needs to know the structure of the distribution of strategies in the population of traders and not only the aggregates quantities. There are no previous studies of this structure in large aggregative games we are aware of. The aim of this thesis is twofold: (i) to investigate the structure of the steady state of some learning dynamics as well as the finite size effects such as first passage times, (ii) to apply those results to help understand the spontaneous emergence of segregation of traders across double auction markets.

Having stated the motivation for this thesis, we give an overview of the methods – mostly coming from statistical physics – that we will use in this thesis. The application of methods from physics to game theory proved to be successful in the past as attested by a substantial body of literature. For example, a number of papers [7, 3, 100, 4] are concerned with finite size effects in birth-death processes in evolutionary game theory. Other studies [86, 88, 87, 42, 76] are concerned with learning dynamics in games. One can also mention [31] where Dall'Asta *et al.* study the "Collaborative Nash equilibria", of public good games on networks. The generalization to random graphs of their model is performed with the well-known cavity method [63]. On top of these works, some recent contributions to mean field game theory originated from statistical physicist [96, 95].

The minority game [28, 24] is also an active area of research in the physics community. As its name suggests, it is a model where players compete to be in the minority. An example of a minority game is "which road to choose to avoid the traffic jam on the way home, knowing that all your colleagues will face the same choice". This means choosing the road the *minority* of your colleagues choose to use. It originates from the El-Farol Bar problem [5] and has received attention from physicists because of its formal relation to spin glasses models. For a non-exhaustive literature review on minority games, we refer to the articles collected in the second part of [24]. Two papers [51, 53] are relevant here as they are concerned with segregation in minority games.

In [51] agents who play a multi-resources minority game are distributed on a graph and only observe their neighbour, whose winning strategy they copy. In such models Huang *et al.* observe a strategy grouping behaviour. The major difference between the model of Huang *et al.* and our model is that the locality of information they assume, which fairly directly leads to segregation, is not a feature of model of segregation in double auction markets. The authors of [53] study *evolutionary minority games*. In such a game, agents have a common memory of the output of previous sessions of the minority game and choose to play either according to it with probability p or against it with probability $1-p$. The agents learn the strategy p and in the long run, a strategy grouping is observed.

In this thesis we will frequently use the Kramers-Moyal expansion [104], Kramers' reaction rate theory [104, 46], mean field theory and also the Freidlin-Wentzell theory. Freidlin-Wentzell generallises Kramers' rate theory to non conservative dynamics [11, 38]. The main book about Freidlin-Wentzell theory being hardly readable for a physicist, we suggest to the interested reader the paper of Bouchet *et al.* [11], which has a very clear introduction to Freidlin-Wentzell theory. The most challenging problem we encountered when using the Freidlin-Wentzell theory is to calculate transition rates between two stable points of a dynamics. This requires finding the minimal action path between those two points. There is a variety of algorithms for this task, two of which are explained in [14, 48]. The ultimate goal of the Freidlin-Wentzell theory is to provide a picture of the behaviour of complex systems in the low noise limit. Such a low noise limit is a recurring theme all along this thesis either because the noise comes from the large size of the system (see Chap. 5) or from the large memory of traders in EWA learning dynamics (see Chap. 2, 3 and 4). As usual in statistical physics in the low noise limit, the problem structure is simplified and often allows for an analytical solution. In the next subsection, we describe the structure of the thesis.

1.3 Organisation of the thesis

In this PhD thesis, we study different learning dynamics in the low noise limit. The connection between the chapters is shown graphically in Fig. 1.1. In Chap. 2 we study the learning dynamics of traders choosing among double auction markets. We first look at the equilibrium of this game in the "Nash sense". Then we compare this Nash equilibrium to the steady state of Experience-Weighted Attraction learning dynamics with fictitious play. We find that in the limit of full fictitious play, large memory and large intensity of choice, the steady states of the learning dynamics do correspond to Nash equilibria. One of the surprising outcomes of our research is that depending on how the relevant limit of Experience-Weighted Attraction learning is approached, the learning dynamics has qualitatively different long-run outcomes. Each corresponds to a Nash equilibrium, and some have groups of players playing with mixed strategies, a scenario not considered before.

In Chap. 2 we focus for simplicity on traders choosing between only two markets. Chap. 3 extends our study to the case of three markets. This connects to a preliminary simulation study of this setting by Alorić [1] and puts the results on a firmer theoretical footing thanks to the minimal action path formalism developed in Chap. 2. This enables us to investigate the existence of strong segregation, where "strong" refers to segregation that persists in the large memory limit.

In real markets, traders have different objectives, depending for example on whether they are speculators or investors. They can have different risk aversion, want to make a profit on different timescales (days, months) etc. To take into account such heterogeneity we study in Chap. 4 a model where the population of traders has heterogeneity in both their memory length and their intensity of choice. In our model, the forgetting rate of traders r represents

the inverse time scale over which agents aim to make a profit. We consider the existence of segregation or otherwise in a population where some traders have an inverse memory $r = 1$ while the rest has a small inverse memory $r \ll 1$. We also study additional heterogeneity in the decision strength β, which is the extent to which traders act on preferences learnt from previously received payoffs.

In Chap. 5 we give an interpretation of reinforcement learning as a pairwise matching process in populations of ideas. This analogy is, on the one hand, motivated by existing work [88, 86, 101, 102, 56] linking reinforcement learning to the replicator equation of evolutionary game theory, with a modified fitness that includes an entropic term. The resulting equation is known as the Sato-Crutchfield equation. On the other hand, there is a large litterature [99, 100, 99, 67] that investigates how one can interpret the original replicator equation as the large size limit of a pairwise matching process and what properties – including e.g. fixation times – this process has for finite population size. Keeping this fact in mind, we show that also the Sato-Crutchfield equation can be written as a pairwise matching process, where the fitness of agents strategies contain an entropic term that accounts for the willingness of players to randomise their strategies, *i.e.* how they set their exploration-exploitation trade-off [94]. We describe birth-death processes in a population of ideas in the limit of a large population as a sequence of simple events (*relaxation* which last a time of order 1, *activation* which takes a time exponential in the size of the population N, ...) which provides a useful intuitive framework for understanding the dynamics of ideas in the large size limit.

Finally, in Chap. 6 we summarise the results of our work and discuss some possible applications as well as future research directions.

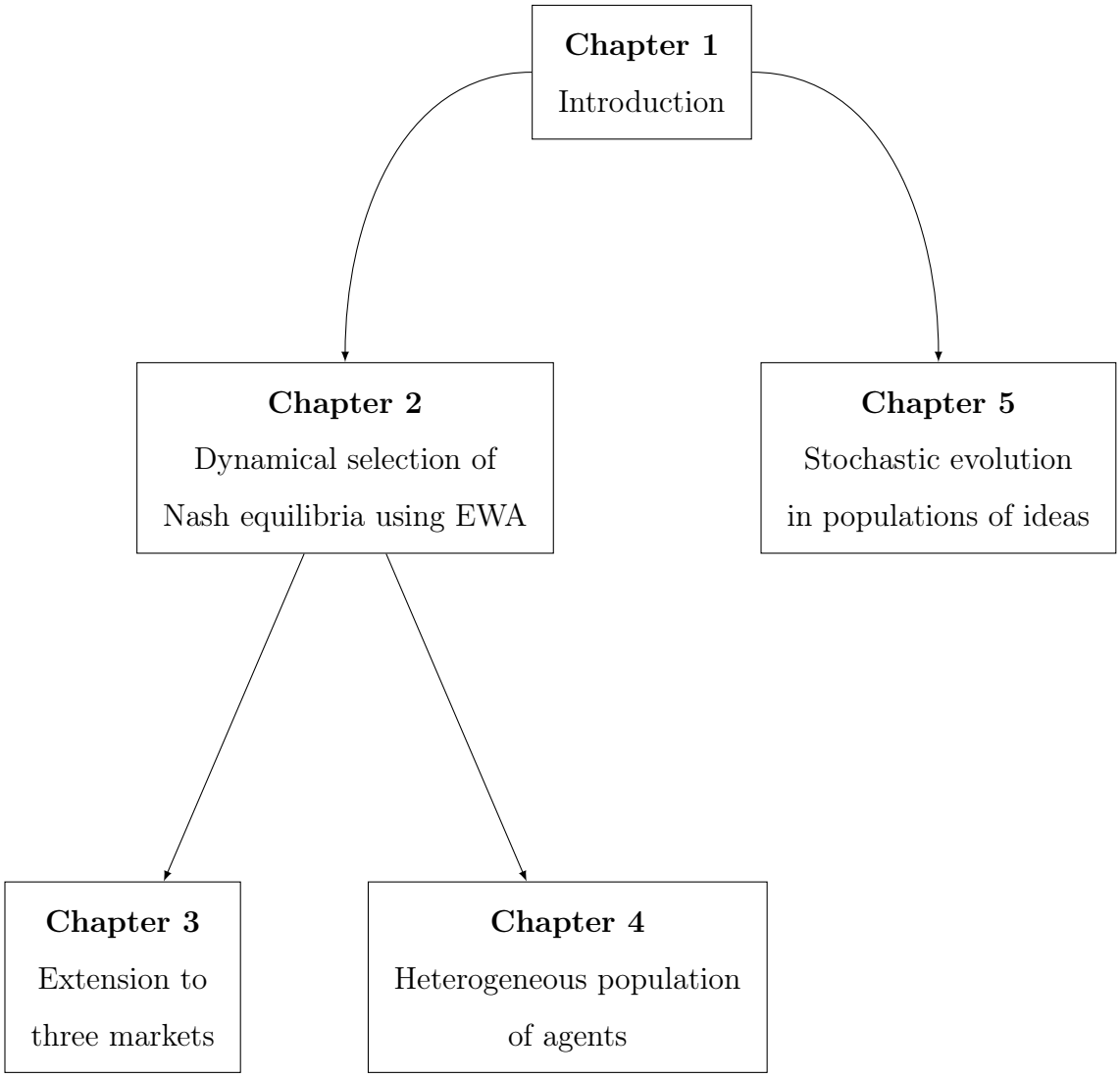

Figure 1.1: Roadmap through the thesis: Chap. 2 is self-contained and describes, among other things, the large deviation formalism used throughout this thesis. Chaps. 3 and 4 rely on the methods developed in Chap. 2 and deal with two extensions of the model of segregation in double auction markets. Chap. 5 is a self-contained study of rare events such as fixation in a *large* population of ideas that follows a birth death process.

CHAPTER 2

Dynamical selection of Nash equilibria using Experience Weighted Attraction Learning

2.1 Introduction

Agent based models describe the dynamics of co-learning and interacting individuals and can be applied in many fields including sociology – with the Schelling model of segregation [89, 84] a famous example – and economics, where the individuals are economic agents. In recent decades, there has been growing interest in the application of agent based models to the study of financial markets; for extensive reviews of such applications we refer to [23, 85]. Among existing models of double auction markets, one can cite the work of Iori et al. [25] and the CAT game [16]. The latter is a market design tournament in which participants were asked to supply automated markets that would perform as well as possible in an economic system populated with automated traders. Spontaneous emergence of preferences for different markets emerged within the population of traders. Unfortunately, the complexity of the CAT

game tournament made it impossible to study this so-called segregation phenomenon by analytical methods, emphasizing the need for a simpler model to understand the phenomenon of segregation. Alorić et al. designed such a minimal version of the CAT game, where traders learn to choose among *two* double auction markets [2]. Also there segregation was observed, as the outcome of the learning dynamics. Whether this result has an interpretation as a game theoretical equilibrium was not addressed, however. This will be one of the two main questions of this chapter: we ask to what extent segregation shows up in the *Nash equilibria* of the game corresponding to the model of Alorić *et al.*. One of the properties of this game are that the payoff agents earn by trading at the different markets depends only on the ratio of the number of buyers and sellers at this market. The game therefore belongs to the class of aggregative games, where payoffs depend on a finite number of macroscopic quantities, called aggregates.

Bearing in mind the above broader context, we consider in this chapter the double auction game of [2] as a paradigmatic example of an aggregative game with an infinitely large number of players. While it is known that finding Nash equilibria in games with a large but finite number of players is computationally hard [32], taking the number of players to infinity can lead to drastic simplifications that make the problem analytically tractable. This is because the limit eliminates some features such as the market impact of the action of a single player [35]. For aggregative games the limit also has convenient mathematical properties: Nash equilibria of infinite games can be characterised as the large size limit of equilibria in games with a finite number of players [21]. An introduction to games with a large number of players would not be complete without mentioning mean field game theory [59, 20, 96, 95], which studies stochastic differential games with an infinite number of players. The underlying formalism here is rather different from the one we use in the rest of this article, however.

Nash equilibria of aggregative game are characterised by the values of the aggregates on which the payoff of any given action depends. To each of these there generally correspond infinitely many different distributions of strategies among the players. In this chapter, the second question we therefore ask is whether and how this degeneracy in the strategy distribution is resolved by the learning dynamics of the corresponding agent based model. This issue of how a Nash equilibrium is selected dynamically has been studied theoretically for games of small size [40] and using numerical simulation for larger games [55, 39, 22], providing results on the speed of convergence and efficiency of certain types of learning dynamics. While these previous studies focused on the value of macroscopic quantities such as the ratio of number of buyers to number of sellers once the learning dynamics has converged, we are interested in going further and investigating the distribution of strategies, which is crucial in order to establish whether there is segregation or not. The specific learning rule we study is Experience Weighted Attraction (EWA) learning, which is well known to reproduce quite accurately the behaviour of human subjects learning to play repeated normal form games [19]. Strategies are encoded by so-called preferences in EWA learning, and the comparison of the *preference distributions* that result from EWA learning dynamics with the properties of the underlying Nash equilibria is one of our main contributions; this is a novel approach that has not to our knowledge been pursued in the existing literature.

Methodically, we argue that in the game we analyse, correspondence with Nash equilibria requires a long memory limit. The EWA dynamics of the agents is then described by a Fokker-Planck equation, and it is the steady states of this that we study. We deploy large deviation methods to detect segregation, where agents split into sub-populations that each play a different strategy. We combine this approach with numerical simulations in order to shed light on the several, qualitatively different, types of preference distribution that can emerge in the steady state of the learning dynamics. These include the two scenar-

ios that are conventionally considered: homogeneous mixed equilibria, where all agents play the same mixed strategy, and heterogeneous pure equilibria, where different agents play different pure strategies [15, 90, 79]. Surprisingly, however, we also find heterogeneous mixed solutions, where the agents play different strategies and these strategies themselves include mixed strategies.

This chapter is organised as follow: in Sec. 2.2 we summarise the minimal model of traders choosing between double auction markets to be studied in the rest of this chapter, as well as the EWA learning dynamics. In Sec. 2.3 we study the Nash equilibria of the aggregative game corresponding to this model, in the limit of a large number of players. In Sec. 2.4 we present a study of the steady states of the learning dynamics in the model of Sec. 2.2 and argue that in the limit of *fictitious play*, *best response dynamics* and *large memory*, these steady states are Nash equilibria. We show that depending on how these multiple limits are approached, the dynamics selects several distinct Nash equilibria, including ones of heterogeneous mixed type. In Sec. 2.5 we present separately the large deviation methods that we use in our study of the steady states of EWA learning in the large memory limit. Sec. 2.6 summarises our results and lays out some avenues for future research. Technical details are relegated to the appendixes.

2.2 Model: Choosing between Double Auction Markets

In this section, we summarise the model of double auction markets of Alorić *et al.* [2]. In this model, a population of co-evolving traders competes to trade by choosing between two double auction markets. This can lead to segregation, where agents spontaneously split into groups with different preferences for the two markets. The model contains three ingredients: (i) the market mechanism

by which the double auction markets process orders to buy and sell, (ii) the way traders set their order prices (this is assumed fixed and not affected by learning) and calculate their payoff, and (iii) the learning procedure that traders use to learn their trading strategy, i.e. their preference for each market. We describe these three ingredients in turn.

Market mechanism

The model assumes that each market processes orders in discrete trading rounds rather than continuously. In each round each trader places at one of the markets an order to buy or sell one unit of the underlying good. An order is denoted (τ,"price") where $\tau \in \{\mathrm{a}, \mathrm{b}\}$ designates the type of order, with a an order to sell (also known as an ask) and b an order to buy (a bid); "price" is the price at which the trader proposes to buy or sell. For example $(\mathrm{b}, 20)$ is an order to buy one unit of good at a price of 20. Once all the traders have sent their orders (see Dynamics of traders), the clearing process begins. The trading price is set by each market using the formula

$$\pi_m = (1 - \theta_m)\langle \mathrm{b} \rangle + \theta_m \langle \mathrm{a} \rangle \tag{2.1}$$

where $\langle \mathrm{b} \rangle$, $\langle \mathrm{a} \rangle$ are the average prices of bids and asks received by the market and θ_m is the bias of market m towards buyers. All the orders on the wrong side of the trading price (*i.e.* an order to buy lower than the trading price or an order to sell higher than the trading price) are rejected. The remaining *valid orders* are *executed* at the trading price by randomly forming pairs of one buyer and one seller until no more pairs can be formed. As the number of valid bids and asks will differ in general, some traders will remain unmatched; they are unable to trade and their orders are not executed.

Order pricing and payoff calculation

As explained above, it is assumed that traders *always* send an order to buy or sell *exactly* one unit of good to only one single market. This is done to keep the model as simple as possible. Following the work of Gode and Sunders [44], traders set the price of their orders with *zero intelligence*: the price of each order to buy (resp. sell) sent by each trader is an independent Gaussian random variable with mean μ_{b} (resp. μ_{a}) and standard deviation $\sigma_{\mathrm{b}} = \sigma_{\mathrm{a}} = 1$. While this assumption may appear drastic at first sight, Gode and Sunders found that traders sending orders to double auction markets with zero intelligence was a good substitute for individual rationality [44]. The model also assumes that each agent chooses randomly whether to buy or sell, with a fixed probability p_{b} that can be different for different agents.

At the end of a trading round, each trader receives as feedback from the market to which they sent their order whether it was executed and if so at which price. From this each trader computes the score of his order $\mathcal{S}$ as either zero, if the order was not executed, or otherwise as the profit of the order. In the model, payoff associated with the order is defined as the absolute value of the difference between order price and trading price. This payoff is random and is affected by: (i) the submitted order price, (ii) the trading price, and (iii) whether the order is executed, which in turn depends on the ratio of number of buyers and sellers in the market where the offer was sent. (We discuss in Sec. 2.3 how the average payoff over these sources of randomness can be calculated in the limit of a large system.)

Dynamics of traders

The remaining part of the behaviour of the traders that the model needs to prescribe is how they learn their respective preferences for the two markets. The assumption is that agents use experience-weighted attraction reinforcement learning (EWA) [19]. They have attractions A_m to each market $m \in \{1, 2\}$, which they update after each trading round n according to

$$A_m(n+1) = \begin{cases} (1-r)A_m(n) + r\mathcal{S}(n) & \text{if market } m \text{ chosen in round } n \\ (1-\alpha r)A_m(n) & \text{otherwise} \end{cases} \tag{2.2}$$

Here $\mathcal{S}(n)$ is the payoff for the order placed at time-step n, α is a *fictitious play parameter* which describes how fast traders decrease the attraction to actions they do not play, and r is the inverse of the agents' memory, defined as the period of time over which they typically remember past payoffs. Based on those attractions $\mathbf{A} = (A_1, A_2)$, traders then randomly choose a market for trade according to the inverse logit or "softmax" function $\sigma_\beta(\cdot)$,

$$\mathbb{P}(\text{trade at market } 1 \mid \mathbf{A}) = \sigma_\beta(A_1 - A_2) = \frac{1}{1 + \exp(-\beta(A_1 - A_2))} \tag{2.3}$$

where β is the intensity of choice that regulates how strongly the agents use the attractions to bias their preferences. A possible extension of this setup, which we do not pursue here, is to allow the traders to learn also their preference for buying and selling, instead of keeping this fixed [2]. In that case there would be four attractions to be learned, for buying and selling at each of the two markets.

We shall use "EWA learning" as a shorthand to designate the above dynamics where traders learn at which market to trade – note that because of this learning process the traders are somewhat more intelligent than the strictly zero-intelligence traders described by Gode and Sunders [44], who in our scenario would choose randomly also where to trade.

In the following we focus largely on a symmetric setup [2], explained in more detail in Sec. 2.3.2 below. There are two classes of agents in this scenario but their distributions of attractions are related by swapping A_1 and A_2 so it is enough to focus on one class. Numerical simulation and theoretical analysis of EWA learning, for $\alpha = 1$, then show that when the intensity of choice β is above a threshold β_c the distribution of the traders' attractions can become bi-modal [2]: the model produces emergent segregation. By way of orientation, example simulation results for β both below and above the segregation threshold are shown in Fig. 2.1.

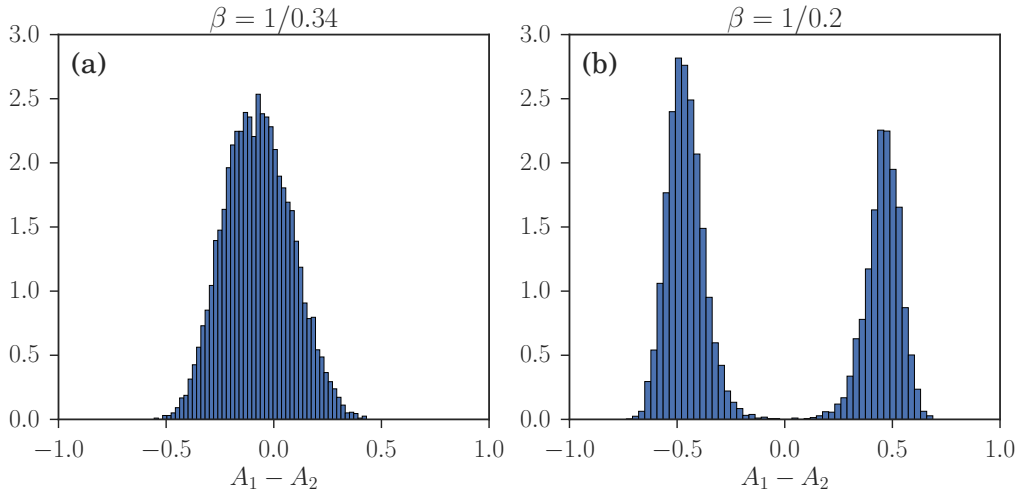

Figure 2.1: Results of a multi-agent simulation of the model of [2] after $5 \cdot 10^4$ rounds of trading among $2 \cdot 10^4$ agents. Parameters for the two markets are $\theta_1 = 1 - \theta_2 = 0.3$, buying preferences for the two classes of agents are $p_{\mathrm{b}}^{(1)} = 1 - p_{\mathrm{b}}^{(2)} = 0.2$, forgetting rate $r = 0.01$ and $\alpha = 1$ (no fictitious play). Shown is the distribution of attraction differences $A_1 - A_2$ across the first group of agents. This is unimodal for intensity of choice β below the segregation threshold as in (a) , but becomes bimodal for larger β: the system shows emergent segregation.

Incomplete versus complete information

One possible cause of heterogeneity in agents' preferences that has been identified in previous studies is incomplete or imperfect information [60]. An obvious question is whether this explains the observation of segregation in the double auction market model described above. Indeed, the agents in this model do have incomplete information about the markets they are trading in: they only receive the stochastic payoffs but do not have access to global information such as the number of buyers and sellers at each market, which they would need in order to estimate their average payoff. As a consequence, traders face the exploration/exploitation dilemma that is common in reinforcement learning [94]. They need to *explore* the whole strategy space (both high and low payoff strategies) to have accurate payoff estimates for their strategies, while at the same time *exploiting* the most profitable strategy by playing it frequently. In the model we consider the trade-off between exploration and exploitation is set by the intensity of choice β [42], with higher values favoring exploitation by making agents choose predominantly the market with the larger attraction.

To address the question of whether segregation is possible also with *perfect* information, we develop in the next section an appropriate game theoretical version of the double auction model discussed above. Once we have determined the Nash equilibria of this game, we will come back to a comparison with the steady state of the EWA learning dynamics, to see how this resolves an indeterminacy in the Nash equilibria.

2.3 Mean field Nash equilibria

We now rephrase the double auction market choice model of Sec. 2.2 in game theoretical language. This will allow us to determine and classify its Nash equilibria in the mean field limit of an infinite number of players. Our aim will be to determine whether in this *perfect information* context there are still signatures of the segregation phenomenon previously found for EWA learning with imperfect information. In Sec. 2.4 we will then see that, in the appropriate limit, the steady states of the EWA learning dynamics are consistent with the Nash equilibria of the model described in this section.

2.3.1 Game theoretical framework

Setting

We consider a population of N traders called players (to be consistent with standard terminology in game theory). Those players are divided into two classes $c \in \{1, 2\}$, of the same size. Each player has fixed buy/sell preferences described by the probability to buy, $p_{\mathrm{b}}^{(c)}$, which depends on his/her class. Each trading round is a round of the game, where each player chooses one of two actions, viz. "send an order to market one" and "send an order to market two"; we label these by $m \in \{1, 2\}$. A *pure strategy* is one where a player always chooses the same action. A *mixed strategy* is one where the player chooses action $m = 1$ with probability $p \in [0, 1]$ and $m = 2$ otherwise. This formalism can be linked to EWA learning as described in Sec. 2.2: there the traders learn which mixed strategy to play, mapping the learned attractions (A_1, A_2) to the probability p using the softmax function $\sigma_\beta(\cdot)$ defined in Eq. (2.3).

Average payoff in a large game

To determine the Nash equilibria, we need to determine the average payoff of a player for a given strategy p, given the (fixed) strategies of all other players. While this calculation would be complicated for finite N, it simplifies in the limit $N \to \infty$ that we consider from now on. Firstly, the trading price at each market becomes non-fluctuating as the average value of bids and asks submitted becomes equal respectively to μ_b and μ_a, up to fluctuations that vanish as $\mathcal{O}(1/\sqrt{N})$.

Secondly, the ratio of the number of buyers and sellers at each market m, which we denote f_m, also becomes non-fluctuating. We can calculate these ratios from the strategy distribution $\phi^{(c)}(p)$ within each class of players, where because of the large N-limit we can neglect the effect of the strategy chosen by any single player to obtain

$$f_1(\phi^{(1)}, \phi^{(2)}) = \frac{p_\mathrm{b}^{(1)}\bar{p}^{(1)} + p_\mathrm{b}^{(2)}\bar{p}^{(2)}}{(1-p_\mathrm{b}^{(1)})\bar{p}^{(1)} + (1-p_\mathrm{b}^{(2)})\bar{p}^{(2)}} \tag{2.4a}$$

$$f_2(\phi^{(1)}, \phi^{(2)}) = \frac{p_\mathrm{b}^{(1)}(1-\bar{p}^{(1)}) + p_\mathrm{b}^{(2)}(1-\bar{p}^{(2)})}{(1-p_\mathrm{b}^{(1)})(1-\bar{p}^{(1)}) + (1-p_\mathrm{b}^{(2)})(1-\bar{p}^{(2)})} \tag{2.4b}$$

Here $\bar{p}^{(c)} = \int \mathrm{d}p\, \phi^{(c)}(p) p$ is the average mixed strategy parameter p in class c. In the above formulas, $N p_\mathrm{b}^{(1)} \bar{p}^{(1)}$ is the typical number of agents of class 1 choosing to buy and to send their buy order to market 1. The relative fluctuations of this number again vanish for $N \to \infty$. The other terms in the expressions for the f_m have analogous interpretations, and the common factor of N cancels.

Based on the above considerations, it becomes a simple matter to calculate the average payoff $\mathcal{P}_{\tau,m}(f_m)$ of buying ($\tau = \mathrm{b}$) or selling ($\tau = \mathrm{a}$) in market m, depending on the market conditions as encoded by f_m (see Appendix 2.A). Our game is therefore *aggregative* [29]: average payoffs are determined only by the *aggregate* quantities f_1 and f_2 that can be calculated from the strategy distribu-

tions $\phi^{(c)}(p)$. Other games in this class include the Cournot oligopoly [33, 30]; in statistical physics language the aggregates would be called order parameters.

In our setup we need to average the payoff $\mathcal{P}_{\tau,m}(f_m)$ further over the probability of buying or selling, giving for a player of class c an average payoff for the action of "going to market m" of

$$\mathcal{P}_m^{(c)}(f_m) = p_{\mathrm{b}}^{(c)}\mathcal{P}_{\mathrm{b},m}(f_m) + (1 - p_{\mathrm{b}}^{(c)})\mathcal{P}_{\mathrm{a},m}(f_m) \tag{2.5}$$

Finally, for a player using a mixed strategy, the resulting payoff $\mathcal{P}^{(c)}(p, f_1, f_2)$ is an average of the payoff at market 1 weighted by p and the payoff at market 2 weighted by $1 - p$:

$$\mathcal{P}^{(c)}(p, f_1, f_2) = p\mathcal{P}_1^{(c)}(f_1) + (1 - p)\mathcal{P}_2^{(c)}(f_2) \tag{2.6}$$

This quantity is the key input into the calculation of the Nash equilibria of our game.

Nash equilibria

We choose to use the following definition of a Nash equilibrium for our game in the limit of an infinite number of players [20]. This definition takes advantage of the fact that we exploited in the payoff calculation, namely that for $N \to \infty$ the aggregate quantities f_1 and f_2 remain constant if a single player changes strategy; in other words, players do not have market impact and their payoff depends only on their own strategy and the *distribution* of the strategies in the population overall.

Definition 1 *Nash equilibrium: The strategy distributions $\phi^{(1)}$ and $\phi^{(2)}$ constitute a Nash equilibrium of the game if the two following conditions are verified:*

$$\text{Support}(\phi^{(1)}) \subseteq \text{argmax}_p \left(\mathcal{P}^{(1)}(p, f_1(\phi^{(1)}, \phi^{(2)}), f_2(\phi^{(1)}, \phi^{(2)}))\right) \tag{2.7a}$$

$$\text{Support}(\phi^{(2)}) \subseteq \text{argmax}_p \left(\mathcal{P}^{(2)}(p, f_1(\phi^{(1)}, \phi^{(2)}), f_2(\phi^{(1)}, \phi^{(2)}))\right) \tag{2.7b}$$

Here the maximization of the payoff on the right hand side is performed over the variable p at constant $\phi^{(c)}$; i.e. each single player maximises their payoff with the aggregate quantities fixed.

In words, the definition means that any strategy that has nonzero probability of being played by a player from class c (*i.e.* in the support of $\phi^{(c)}$) must maximise the player's payoff. We will now apply this definition to determine the different classes of Nash equilibria that exist in the double auction market choice game.

2.3.2 Classification of Nash equilibria

Equal payoff constraints

We will classify Nash equilibria according to two characteristics. If all agents in a class play the same strategy $p = \bar{p}^{(c)}$, the distribution $\phi^{(c)}(p)$ is a delta-distribution $\delta(p - \bar{p}^{(c)})$ and we call the equilibrium *homogeneous* for that class, otherwise—when different players in the same class use different p— we refer to the equilibrium as *heterogeneous*. The second characteristic is the strategy type: if all agents in a class play the pure strategies $p = 0$ or $p = 1$ we call the equilibrium *pure*, otherwise *mixed*. Combining these two characteristics then divides equilibria for each class into four possible types.

To obtain a classification of the possible overall Nash equilibria, note that the function being maximised in Eq. (2.7a,2.7b), viz. $p \to \mathcal{P}^{(c)}(p, f_1(\phi^{(c)}, \phi^{(2)}), f_2(\phi^{(1)}, \phi^{(2)}))$ is *linear* in p. As a consequence, if it is

not constant, it has a single maximum on one of the boundaries of the interval $[0, 1]$ where it is defined. A glance at (2.6) shows that the payoff function is constant if and only if $\phi^{(1)}$ and $\phi^{(2)}$ are such that the payoffs at the two markets are equal:

$$\mathcal{P}_1^{(c)}\left(f_1(\phi^{(1)}, \phi^{(2)})\right) = \mathcal{P}_2^{(c)}\left(f_2(\phi^{(1)}, \phi^{(2)})\right) \tag{2.8}$$

If (and only if) this *equal payoff condition* is satisfied, the strategy distribution $\phi^{(c)}(p)$ can be nonzero for any $p \in [0, 1]$. This can be interpreted by saying that, if in a class there are players that go to the first and the second market, the only way for none of them to have an incentive to move to another market is for the payoff at the two markets to be the same.

If the equal payoff condition is not met for a class, we have to have either

$$\mathcal{P}_1^{(c)}\left(f_1(\phi^{(1)}, \phi^{(2)})\right) > \mathcal{P}_2^{(c)}\left(f_2(\phi^{(1)}, \phi^{(2)})\right), \qquad \phi^{(c)}(p) = \delta(p-1), \qquad \bar{p}^{(c)} = 1 \tag{2.9}$$

or

$$\mathcal{P}_1^{(c)}\left(f_1(\phi^{(1)}, \phi^{(2)})\right) < \mathcal{P}_2^{(c)}\left(f_2(\phi^{(1)}, \phi^{(2)})\right), \qquad \phi^{(c)}(p) = \delta(p), \qquad \bar{p}^{(c)} = 0 \tag{2.10}$$

In both cases the strategy distribution is homogeneous pure, and the entire class of agents goes to the market with the higher payoff.

Types of Nash equilibria

We can now proceed to find the possible types of overall Nash equilibria for our game. Because f_1 and f_2 are fixed once $\bar{p}^{(1)}$ and $\bar{p}^{(2)}$ are known, the equal payoff condition for each class defines a line of points in the $(\bar{p}^{(1)}, \bar{p}^{(2)})$ plane. This line can consist of several distinct pieces as shown in the examples in Fig. 2.2, where equal payoff lines are plotted for both class $c = 1$ (full lines) and $c = 2$ (dashed lines).

The discussion above can now be summarised in graphical terms as follows: a point in the $(\bar{p}^{(1)}, \bar{p}^{(2)})$-plane is a Nash equilibrium if for each class the point is either on the equal payoff line, or on the boundary (specified by $\bar{p}^{(c)} = 1$ or $= 0$) corresponding to the market where the class has the higher payoff. Combining these options for the two classes, the first and for our purposes most interesting type of Nash equilibrium that results is a point at an intersection of two equal payoff lines, away from the boundaries. We call such a point a *potentially heterogeneous* Nash equilibrium. Here both $\bar{p}^{(1)}$ and $\bar{p}^{(2)}$ are strictly between 0 and 1. The strategy distributions can then be either

- homogeneous mixed, with $\phi^{(c)} = \delta(p - \bar{p}^{(c)})$, or
- heterogeneous pure, with $\phi^{(c)} = (1 - \bar{p}^{(c)})\delta(p) + \bar{p}^{(c)}\delta(p - 1)$, or
- heterogeneous mixed otherwise.

These three different cases are illustrated schematically in Fig. 2.3. The homogeneous mixed case can be viewed as the Nash equilibrium analogue of the unimodal distribution in the stochastic simulations shown in Fig. 2.3; in the heterogeneous mixed case the strategy distribution is arbitrary except for its fixed mean $\bar{p}^{(c)}$. The fact that the Nash equilibrium conditions here allow both homogeneous and heterogeneous strategy distributions motivates our use of the term "potentially heterogeneous". It also shows that one needs dynamical information to say more about the strategy distribution shapes, as explored in detail in Sec. 2.4.

A second type of Nash equilibrium results when the equal payoff condition is obeyed for only one class while the other class is at a boundary. We then speak of a *partially potentially heterogeneous* Nash equilibrium, because one class of players has a homogeneous pure strategy distribution while the other strategy distribution is of one of the three types listed in the bullet points above.

Finally, Nash equilibria unconstrained by either of the equal payoff conditions must be in on of the four corners of the square $(\bar{p}^{(1)}, \bar{p}^{(2)}) \in [0,1]^2$; we call them *homogeneous pure* equilibria as the strategy distributions for both classes are then of this type. These equilibria can be further subdivided depending on whether both classes go to the same market or not. The former type always exists as if one of the traders tries to trade in the empty market s/he will earn a payoff of 0 which is smaller than the payoff s/he could earn in the non-empty market. In the latter type, each market is used only by traders of one class, who trade with each other there.

Plots in the $(\bar{p}^{(1)}, \bar{p}^{(2)})$-plane as shown in Fig. 2.2 are a convenient graphical tool to assess the existence of potentially heterogeneous, potentially partially heterogeneous and homogeneous pure Nash equilibria. Potentially heterogeneous equilibria are found directly as interior crossing points of the equal payoff curves for the two classes. A partially heterogeneous Nash equilibrium corresponds to a point (see Fig. 2.2(b)) that is located at the intersection of the equal payoff curve of class 1 (resp. 2) and a horizontal (resp. vertical) boundary. This criterion identifies a list of (usually four) candidate equilibria. To have an actual equilibrium the payoffs of the markets for the homogeneous pure class need to have the correct order, e.g. for a candidate point located on the axis $\bar{p}^{(2)} = 1$, the payoff at market 1 has to be higher for class 2 players than the payoff at market 2. By drawing arrows indicating payoff ordering as explained in the caption of Fig. 2.2, this can be summarised by saying that the arrows must point *towards* the boundary that a candidate point for a potentially partially heterogeneous Nash equilibrium lies on. In Fig. 2.2, this leaves two equilibria of this type as marked by the red circles.

Finally, for a heterogeneous pure Nash equilibrium where the two classes of players choose different markets, the two candidate points are the top left or bottom right corner. These are again Nash equilibria provided they have the correct ordering of payoffs, which requires that the arrows drawn in the figure

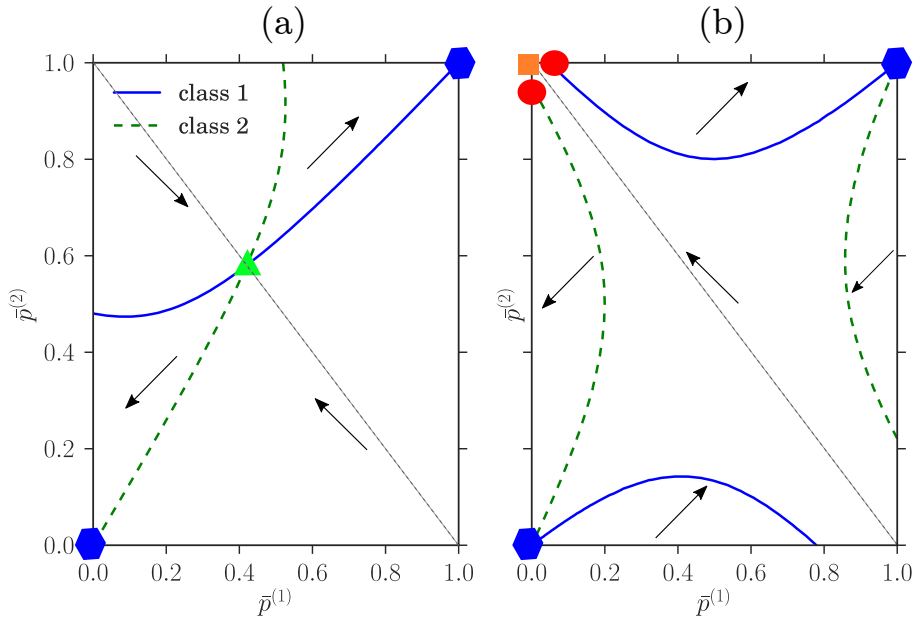

Figure 2.2: Values of $\bar{p}^{(1)}, \bar{p}^{(2)}$ for which the equal payoff constraints are verified for class $c = 1$ (blue, solid) and class $c = 2$ (green, dashed). The arrows point to $(s^{(1)}, s^{(2)})$ where $s^{(c)} \in \{0, 1\}$ indicates the profit-maximizing strategy of traders from class c, in each distinct area of the plane. In panel (a) where $\theta_1 = 1 - \theta_2 = 0.3$, $p_{\mathrm{b}}^{(1)} = 1 - p_{\mathrm{b}}^{(2)} = 0.2$, there exists a heterogeneous equilibrium (green triangle), located at the intersection of the two equal payoff curves. In panel (b), $\theta_1 = 1 - \theta_2 = 0.2$, $p_{\mathrm{b}}^{(1)} = 1 - p_{\mathrm{b}}^{(2)} = 0.45$, and the equal payoff curves do not cross. There is then no potentially heterogeneous Nash equilibrium, but the direction of the arrows shows that a homogeneous pure equilibrium (orange square) with the two classes going to different markets exists. There are also two partially heterogeneous Nash equilibria (red circles, see main text). In both (a) and (b) there exist homogeneous pure Nash equilibria where the whole population trades at the same market (blue hexagons). The dotted line indicates the location of the symmetric equilibria that we mostly focus on.

point towards this corner. In Fig. 2.2(b) this is the case for the top left corner (orange square).

We can now look at how the existence of the different types of Nash equilibria depends on the system parameters, which are the market biases θ_m and the buying preferences $p_{\mathrm{b}}^{(c)}$. We follow Ref. [2] in focusing on a symmetric setup where the two markets have opposite biases in favour of buyers and sellers. As $\theta = 0.5$ corresponds to the absence any bias, this means $\theta_1 + \theta_2 = 1$. Similarly we assume that the players fall into two symmetric groups with respect to their buying preferences, with those in class 1 preferring to buy ($p_{\mathrm{b}}^{(1)} < 0.5$) and the others having the opposite preference $p_{\mathrm{b}}^{(2)} = 1 - p_{\mathrm{b}}^{(1)}$. With these choices, we can show in Fig. 2.4 the regions where the different types of Nash equilibria exist as a function of $p_{\mathrm{b}}^{(1)}$ and θ_1. It turns out that the two examples shown in Fig. 2.2 cover the two generic cases: in addition to homogeneous pure Nash equilibria where both classes go to the same market, which always exist, one has either a potentially heterogeneous Nash equilibrium as in Fig. 2.2(a), or a homogeneous pure equilibrium with the two classes at different markets and two potentially partially heterogeneous equilibria (Fig. 2.2(b)). These two cases are mutually exclusive. An analytical expression for the boundary between the zones where they exist can also be obtained as detailed in Appendix 2.B.

Returning to the broader picture, the Nash equilibrium analysis of the double auction market choice game clearly shows that there is *potential* for segregation: as illustrated in Fig. 2.3, heterogeneous pure strategy distributions have two peaks that indicate players within a class separating into two distinct subpopulations playing opposite pure strategies. Heterogeneous mixed strategies can similarly have two or more peaks. This emergence of segregation shows that the observations of segregation in a previous study of EWA learning [2] were not based on purely dynamical effects. We also find qualitatively similar trends, e.g. the equilibria where both classes of players can be segregated (potentially heterogeneous) are most prevalent in Fig. 2.4 when the two markets are identical ($\theta_1 = 0.5$), showing that segregation is not a trivial consequence of differences between markets.

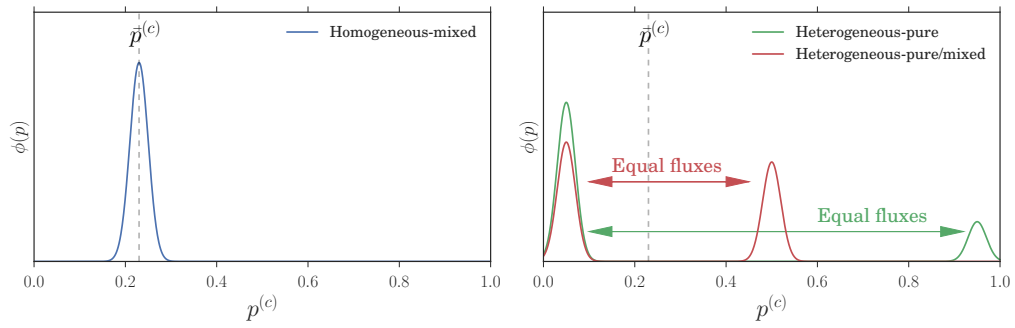

Figure 2.3: Schematic representation of three different types of strategy distribution $\phi(p)$ that all have the same mean $\bar{p}$ (dashed line): homogeneous mixed distribution (left panel), heterogeneous mixed (red curve, right panel) heterogeneous pure (green curve, right panel). Peaks in the distribution are shown broadened as they would be in EWA learning at finite decision strength β; as Nash equilibria they would become sharp (delta-distributions). The right panel illustrates that, when a strategy distribution has two distinct peaks, it can represent a steady state of the learning dynamics only when the fluxes of agents moving from one peak to the other balance in the two directions (see Sec. 2.5).

However, the Nash equilibrium conditions only identify the means of the strategy distributions $\phi^{(1)}$ and $\phi^{(2)}$. As we saw, this means for a potentially heterogeneous (or potentially partially heterogeneous) equilibrium that we cannot decide whether the underlying strategy distribution is homogeneous (mixed) or heterogeneous, nor do we know whether a heterogeneous mixed strategy distribution would actually have two distinct peaks as required for the concept of segregation to make sense. We therefore study next under what conditions EWA learning *dynamics* as defined in Sec. 2.2 reaches as its steady state a Nash equilibrium of our system. Once this connection is established, we ask which particular Nash equilibria are selected as possible steady states of EWA learning. Put differently, does the learning dynamics break the indeterminacy of the Nash equilibrium conditions?

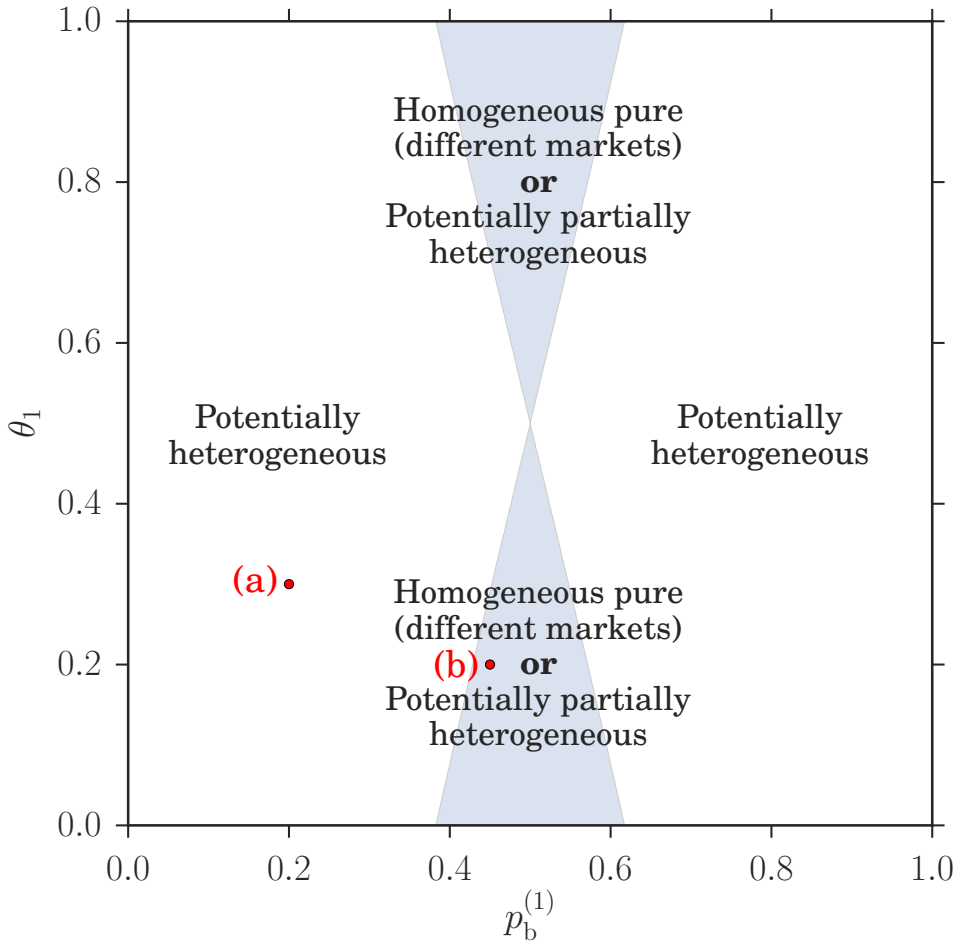

Figure 2.4: Phase diagram for existence of different types of Nash equilibria for a system with symmetric price setting parameters $\theta_1 = 1 - \theta_2$ and buying preferences $p_b^{(1)} = 1 - p_b^{(2)}$. The types of equilibria in this plot are explained in Sec. 2.3.2 and a graphical method to check their existence is shown in Fig. 2.2. The labels (a) and (b) correspond to the panels there. Note that the two homogeneous pure Nash equilibria where both classes of player trade at the same market are not shown as they exist everywhere.

2.4 EWA learning in double auction markets

In this section, we study the steady states of the EWA learning dynamics defined in Sec. 2.2 in a game with a large number of players. We are interested in particular when different types of steady state strategy distributions, as sketched in Fig. 2.3, can occur.

We argue in Sec. 2.4.1 that one expects the steady state of the EWA learning dynamics to approach a Nash equilibrium of the model described in Sec. 2.3 in the joint limit where the fictitious play coefficient $\alpha \to 0$, the intensity of choice $\beta \to \infty$ and the inverse memory length $r \to 0$. In principle our task is thus to find the steady state of EWA learning and then to take this joint limit. It turns out, however, that this is far from trivial. The reason is shown by the phase diagram in Fig. 2.5, where the limit $r \to 0$ has already been taken. What is notable is that there are different regions in the phase diagram where the steady state strategy distributions are homogeneous and heterogeneous, respectively. The Nash equilibrium limit point $(\alpha, 1/\beta) = (0, 0)$ can be approached along paths within either of these regions, which means there will be several possible limiting strategy distributions of EWA learning, and it is these that we will want to identify. Note that we focus generally on system parameters where potentially heterogeneous Nash equilibria exist (see Fig. 2.4), for which the EWA learning phase diagram has the generic structure of Fig 2.5.

We introduce in Sec. 2.4.2 the Kramers-Moyal expansion for the EWA learning dynamics on which the rest of the analysis is based. In particular, we study homogeneous and heterogeneous distributions of preferences in Sec. 2.4.4 and Sec. 2.4.5, respectively, and analyse how they approach Nash equilibria in the relevant limit. The large deviation methods we deploy for the heterogeneous case are described separately in Sec. 2.5.

As before we choose to concentrate on settings with symmetric market biases and buy/sell preferences, and within those on steady states of the EWA learning dynamics that also have symmetric aggregates $\bar{p}^{(1)} = 1 - \bar{p}^{(2)}$. This captures the dominant steady states, simplifies the numerical analysis (see Sec. 2.5) and also makes it easier to illustrate the concepts. In the graphical representation of Fig. 2.2, the steady states we are considering lie on the diagonal from top left to bottom right (dotted line).

2.4.1 Nash Equilibria as limits of EWA learning

In the game theoretical study of Sec. 2.3, we considered a large game ($N \to \infty$). The Nash equilibria we studied assume implicitly (i) that each player is able to evaluate his expected payoff (*full information assumption*), (ii) that this evaluation averages appropriately over all stochastic effects (*no fluctuation assumptions*) and (iii) that the players always choose the action with the highest payoff (*best response assumption*). One therefore expects a learning dynamics that verifies these same assumptions to converge to one of the Nash equilibria we characterised in Sec. 2.3.

We now consider when the above assumptions hold for EWA learning dynamics. If we want the players' attractions to be accurate estimates of the payoffs for the corresponding action (assumption (i)) we require $\alpha \to 0$ to ensure that the attractions to actions that are not played do not decrease over time. To average over payoff fluctuations (assumption (ii)) we further need to work in the large memory limit $r \to 0$. To see this, note that in each training round the players' attractions are modified only by an amount of order r. For small r, attractions therefore change substantially after $\sim 1/r$ training rounds. This means the players effectively average the payoffs over many trading rounds that take place while their attractions and hence their strategies remain fixed, and in the limit obtain the correct expected payoffs [88]. Finally, a large inten-

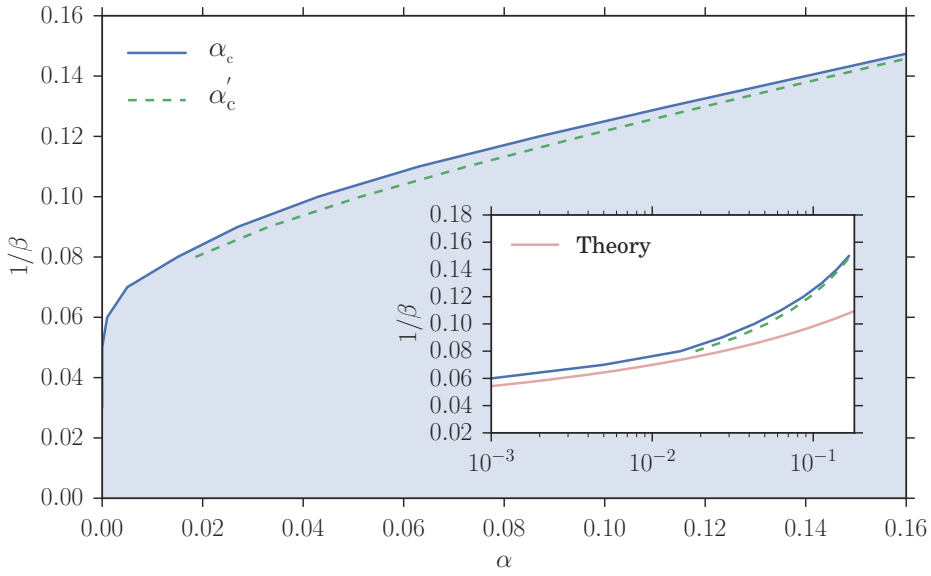

Figure 2.5: EWA learning phase diagram. The blue zone shows the region of the $(\alpha, 1/\beta)$-plane where the steady state strategy distribution of each of the two classes of agents is heterogeneous. Elsewhere, including in particular on the line $\alpha = 0$, the strategy distribution is homogeneous. The blue line shows the threshold α_c where the distribution switches from homogeneous to heterogeneous mixed. As α is increased further beyond a threshold α_c' (dashed green line), the strategy distribution becomes heterogeneous pure. The market and trader parameters for this diagram are $\theta_1 = 1 - \theta_2 = 0.3$ and $p_b^{(1)} = 1 - p_b^{(2)} = 0.2$. Inset: Threshold curves plotted with a logarithmic α-axis. The red line shows the exponential dependence of the characteristic values of α on β (with an arbitrary prefactor) that is expected from the theoretical considerations in Appendix 2.D.

sity of choice ($\beta \to \infty$) ensures that players best respond to their attractions, so that EWA learning in that limit also verifies assumption (iii).

2.4.2 Kramers-Moyal expansion for $r \to 0$

Of the three limits identified above we take first the large memory limit $r \to 0$. In this limit—and the large system limit $N \to \infty$, which we always assume—the dynamics of EWA learning can be described by a (nonlinear) Fokker-Planck equation [2]. This is derived by a Kramers-Moyal expansion truncated at the second order; we defer the details to Appendix 2.C. Denoting by $\mathbb{P}(\mathbf{A}^{(c)}, t)$ the distribution of attractions of traders from class c, where $\mathbf{A}^{(c)} = (A_1^{(c)}, A_2^{(c)})$ is a vector gathering the attractions towards market 1 and 2, the Fokker-Planck equation describing the time evolution of this distribution is

$$\begin{aligned} \partial_t \mathbb{P}(\mathbf{A}^{(c)}, t) &= -\sum_{m=1}^{2} \partial_{A_m^{(c)}} \left[\mu_m^{(c)}(\mathbf{A}^{(c)}, \bar{p}^{(1)}, \bar{p}^{(2)}) \mathbb{P}(\mathbf{A}^{(c)}, t) \right] \\ &\quad + \frac{r}{2} \sum_{m,m'=1}^{2} \partial_{A_m^{(c)}} \partial_{A_{m'}^{(c)}} \left[\Sigma_{mm'}^{(c)}(\mathbf{A}^{(c)}, \bar{p}^{(1)}, \bar{p}^{(2)}) \mathbb{P}(\mathbf{A}^{(c)}, t) \right] \end{aligned} \tag{2.11}$$

Here time $t = rn$ is a rescaled version of the number of trading rounds n, while $\bar{p}^{(1)}$ and $\bar{p}^{(2)}$ are the average fractions of traders from class 1 (resp. class 2) choosing to go to the first market. These fractions are obtained simply by averaging the probability of choosing market 1 as defined in (2.3) over the relevant distribution of attractions:

$$\bar{p}^{(c)} = \int \mathrm{d}\mathbf{A}^{(c)}\, \mathbb{P}(\mathbf{A}^{(c)}, t) \sigma_\beta(A_1^{(c)} - A_2^{(c)}) \tag{2.12}$$

Formally, $\bar{p}^{(1)}$ and $\bar{p}^{(2)}$ are therefore functionals of the probability distributions $\mathbb{P}(\mathbf{A}^{(c)}, t)$ It is this dependence that makes the Fokker-Planck equation nonlinear, and couples the dynamics of the attraction distributions in class 1 and 2.

At fixed values of $\bar{p}^{(1)}$ and $\bar{p}^{(2)}$, the Fokker-Planck equation (2.11) describes for each class the Langevin dynamics of the attraction vector $\mathbf{A}^{(c)}$ of a *single agent*, with deterministic drift vector $\mu_m^{(c)}$ and (multiplicative) white noise with

covariance matrix $r\Sigma^{(c)}_{mm'}$. The form of the drift follows directly from the original EWA dynamics (see Appendix 2.C)

$$\mu_1^{(c)}(\mathbf{A}^{(c)}, \bar{p}^{(1)}, \bar{p}^{(2)}) = \left[\mathcal{P}_1^{(c)}(f_1(\bar{p}^{(1)}, \bar{p}^{(2)})) - A_1^{(c)}\right] \sigma_\beta(A_1^{(c)} - A_2^{(c)}) \\ - \alpha A_1^{(c)} \left[1 - \sigma_\beta(A_1^{(c)} - A_2^{(c)})\right] \tag{2.13}$$

The first term describes the change in the attraction to market 1 (in square brackets), weighted with the probability of the agent choosing that market. The second term corresponds to the opposite case where the agent chooses market 2.

The Fokker-Planck equation (2.11) is of course impossible to solve in closed form in general. A special case is the limit $r \to 0$, assuming the population is initially homogeneous, *i.e.* a delta-distribution. Homogeneity is then maintained over time for $r = 0$, where the dynamics are deterministic, and Eq. (2.11) gives for the time evolution of the locations of the peaks of the attraction distributions the equations

$$\partial_t A_m^{(c)} = \mu_m^{(c)}(\mathbf{A}^{(c)}, \bar{p}^{(1)}, \bar{p}^{(2)}) \tag{2.14}$$

Together with

$$\bar{p}^{(c)}(t) = \sigma_\beta(A_1^{(c)}(t) - A_2^{(c)}(t)) \tag{2.15}$$

one then has a system of nonlinear differential equations that is straightforward to solve numerically. We call this the *homogeneous populations dynamics*, where the population changes over time but remains homogeneous.

For nonzero r, analysing the EWA Fokker-Planck equation becomes more difficult because the attraction distributions broadens and can indeed develop multiple peaks. As we are primarily interested in long-time steady states, we focus on this somewhat simpler case. The task at hand here is a self-consistency problem: find a set of aggregates $\bar{p}^{(1)}$, $\bar{p}^{(2)}$ for which the steady state solution of the Fokker-Planck equation, when inserted into (2.12), gives back the original aggregates. If we call $\tilde{p}^{(c)}(\bar{p}^{(1)}, \bar{p}^{(2)})$ the aggregates calculated from the steady

state solution, the self-consistency equations are simply $\tilde{p}^{(c)}(\bar{p}^{(1)}, \bar{p}^{(2)}) = \bar{p}^{(c)}$.

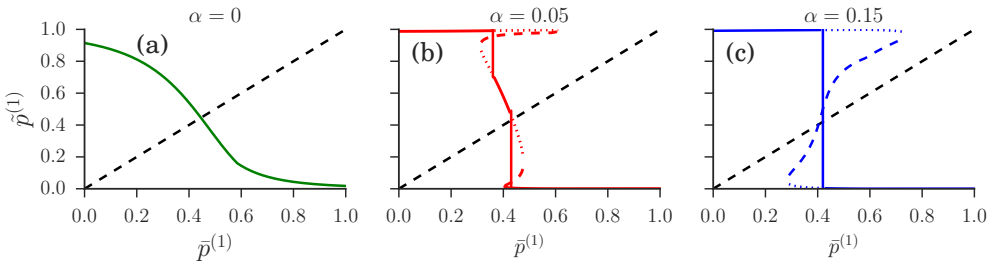

Figure 2.6: New aggregate $\tilde{p}^{(1)}$ calculated from steady state of single agent dynamics at "old" aggregate value $\bar{p}^{(1)}$ (for $r \to 0$). Steady states are peaked around stable fixed points (solid/dotted), which are connected by unstable fixed points (dashed). In (a) only one such peak exists for any $\bar{p}^{(1)}$. The physical steady state is found from the self-consistency requirement $\tilde{p}^{(1)} = \bar{p}^{(1)}$ (dot-dashed line). In (b,c) there are steady states with up to three peaks, but generically all but one have a weight exponentially suppressed in $1/r$ so that $\tilde{p}^{(1)}(\bar{p}^{(1)})$ (solid line) follows the curve for a single fixed point. At specific aggregate values the dominant peak switches and two peaks can coexist (vertical solid lines). In (b) there are two such transitions; in (c) the middle fixed point from (b) has disappeared and there is only one transition, between branches of $\tilde{p}^{(1)}$ that are close to 0 and 1. In (b,c) the intersection with the diagonal is *at* a switch, giving a heterogeneous steady state with two peaks of comparable weight. Market and trader parameters for this figure are as in Fig. 2.5; intensity of choice $\beta = 1/0.1$.

2.4.3 Steady state of EWA Fokker-Planck equation

The remaining challenge is now to determine, for small r, the steady state solution of the Fokker-Planck equation for given aggregates $\bar{p}^{(1)}, \bar{p}^{(2)}$. As explained above, we can think of this as the steady state distribution for the dynamics of

a single agent, given a fixed state of the population. In the limit $r \to 0$ those dynamics are almost deterministic so that the agent will spend almost all of her/his time near the stable fixed points of the drift $\mu_m^{(c)}$. Accordingly, $\mathbb{P}(\mathbf{A}^{(c)})$ will be peaked near these points, with the peak width being of the order of the standard deviation of the Langevin noise, *i.e.* $O(\sqrt{r})$.

For aggregate values where there is only one stable single agent fixed point, $\mathbb{P}(\mathbf{A}^{(c)})$ becomes a delta-distribution centred at that point for $r \to 0$, so we have a steady state with a homogeneous distribution of attractions and hence strategies. The self-consistency condition for such a steady state is then simply the stationarity condition for the homogeneous population dynamics (2.14) together with (2.15). The graphical solution of this condition is illustrated in Fig. 2.6(a).

When there are multiple stable single agent fixed points, $\mathbb{P}(\mathbf{A}^{(c)})$ for $r \to 0$ will become a sum of delta-distributions at these points. The remaining task is then to find the *weight* of each of these peaks. We explain how to use *large deviation* methods for this purpose in Sec. 2.5. The idea is that the peak weights are determined by the balance of fluxes of agents transitioning from one peak to another. For small r, the dominant r-dependence of these fluxes comes from exponential factors of the form $\exp(-\mathcal{S}/r)$. Fluxes can then balance for $r \to 0$ only when the "action" $\mathcal{S}$, which represents an effective activation barrier, is the same for the transition from one peak to the other as for the reverse transition. This condition, which is represented schematically in Fig. 2.3, allows one to determine the aggregate values where multiple peaks can coexist in $\mathbb{P}(\mathbf{A})$. At these aggregate values the steady state solution switches between two single peaked solutions. This switch happens within an aggregate value range of $O(r)$ that vanishes as $r \to 0$, giving vertical sections in the plot of $\tilde{p}^{(c)}$ versus $\bar{p}^{(c)}$ as shown in Fig. 2.6(b). If the intersection with the diagonal $\tilde{p}^{(c)} = \bar{p}^{(c)}$ occurs in one of these vertical sections, as in the example in Fig. 2.6(b), the actual peak weights can be determined indirectly from the fact that the appropriate

weighted combination of the $\tilde{p}^{(c)}$ from the single peaks must give $\bar{p}^{(c)}$. Note that one can show generally (see Appendix 2.D) that in each agent class there can be at most three stable fixed points, so that each $\mathbb{P}(\mathbf{A}^{(c)})$ can have at most three peaks. By choosing an appropriate aggregate value, at most two of these peaks can be made to have finite weight for $r \to 0$. Obtaining three peaks with finite weight requires one to tune α to α_c' at given β, giving the dashed green phase boundary in Fig. 2.5. Intuitively, at α_c' the two transitions in Fig. 2.6(b) have moved horizontally so that they occur at the same aggregate value.

We will next study the homogeneous steady states of EWA learning dynamics. Given the structure of the phase diagram that we anticipated in Fig. 2.5, the easiest way to ensure that steady states are homogeneous in the Nash equilibrium limit is to take $\alpha = 0$.

2.4.4 Homogeneous attraction distributions

Kramers-Moyal expansion for $\alpha = 0$

We saw above that the dynamics of a homogeneous distributions of agents within each class is described, for $r \to 0$ by (2.14,2.15). In steady state the right-hand side of (2.14) needs to vanish, hence using $\alpha = 0$ in (2.13) and its analogue for $m = 2$ one has

$$0 = [\mathcal{P}_1^{(c)}(f_1(\bar{p}^{(1)}, \bar{p}^{(2)})) - A_1^{(c)}]\sigma_\beta(A_1^{(c)} - A_2^{(c)}) \tag{2.16}$$

$$0 = [\mathcal{P}_2^{(c)}(f_2(\bar{p}^{(1)}, \bar{p}^{(2)})) - A_2^{(c)}]\sigma_\beta(A_2^{(c)} - A_1^{(c)}) \tag{2.17}$$

Here the aggregates on which f_1 and f_2 depend are given by $\bar{p}^{(c)} = \sigma_\beta(A_1^{(c)} - A_2^{(c)})$. In (2.17), $\sigma_\beta(A_1^{(c)} - A_2^{(c)})$ cannot vanish at any finite β, so the condition for a homogeneous state is simply

$$\mathcal{P}_m^{(c)}(f_1(\bar{p}^{(1)}, \bar{p}^{(2)})) - A_m^{(c)} = 0 \tag{2.18}$$

which needs to be verified for each market m and each class c. This means that for each player, in the steady state of the EWA learning dynamics, the respective attraction to each market equals the expected payoff there. The aggregates calculated from the steady state are therefore

$$\tilde{p}^{(c)}(\bar{p}^{(1)}, \bar{p}^{(2)}) = \sigma_\beta \left(\mathcal{P}_1^{(c)}(f_1(\bar{p}^{(1)}, \bar{p}^{(2)})) - \mathcal{P}_2^{(c)}(f_2(\bar{p}^{(1)}, \bar{p}^{(2)})) \right) \tag{2.19}$$

We now need to solve the self-consistency condition $\tilde{p}^{(c)} = \bar{p}^{(c)}$as explained in Sec. 2.4.3. This can be visualised most easily if we focus on symmetric situations where $\bar{p}^{(1)} = 1 - \bar{p}^{(2)}$: one just has to plot the curve $\sigma_\beta(\mathcal{P}_1^{(1)} - \mathcal{P}_2^{(1)})$ vs $\bar{p}^{(1)}$ and intersect it with the diagonal, as shown in Fig. 2.6(a).

To retrieve EWA steady states corresponding to Nash equilibria, we need to consider the limit $\beta \to \infty$ of high intensity of choice. Then $\sigma_\beta(\mathcal{P}_1^{(1)} - \mathcal{P}_2^{(1)})$ approaches one if the payoff at the first market $\mathcal{P}_1^{(1)}$ is larger than at the second, otherwise zero. Where the payoffs are equal, a step in the curve results, which will always produce an intersection and hence a self-consistent solution. Because of the payoff equality, such solutions correspond exactly to potentially heterogeneous Nash equilibria (see Eq. (2.8) in Sec. 2.3). Here this type of Nash equilibrium is realised in a *homogeneous mixed* form: all players from class 1 play the same strategy, choosing market 1 with probability $\bar{p}^{(1)}$.

If the payoffs $\mathcal{P}_1^{(1)}$ and $\mathcal{P}_2^{(1)}$) are different across the entire range of $\bar{p}^{(1)}$, we have a different scenario: assuming $\mathcal{P}_1^{(1)} > \mathcal{P}_2^{(1)}$ for definiteness, $\sigma_\beta(\mathcal{P}_1^{(1)} - \mathcal{P}_2^{(1)})$ tends to one for $\beta \to \infty$, hence the only self-consistent solution is $\bar{p}^{(1)} = 1$.

This corresponds to a *homogeneous pure* Nash equilibrium, with—because of the assumed symmetry—the two classes of players trading at different markets.

To show the approach to the large β-limit, we show in Fig. 2.7 numerically determined values of $\bar{p}^{(1)}$, the fraction of traders from the first class going to the

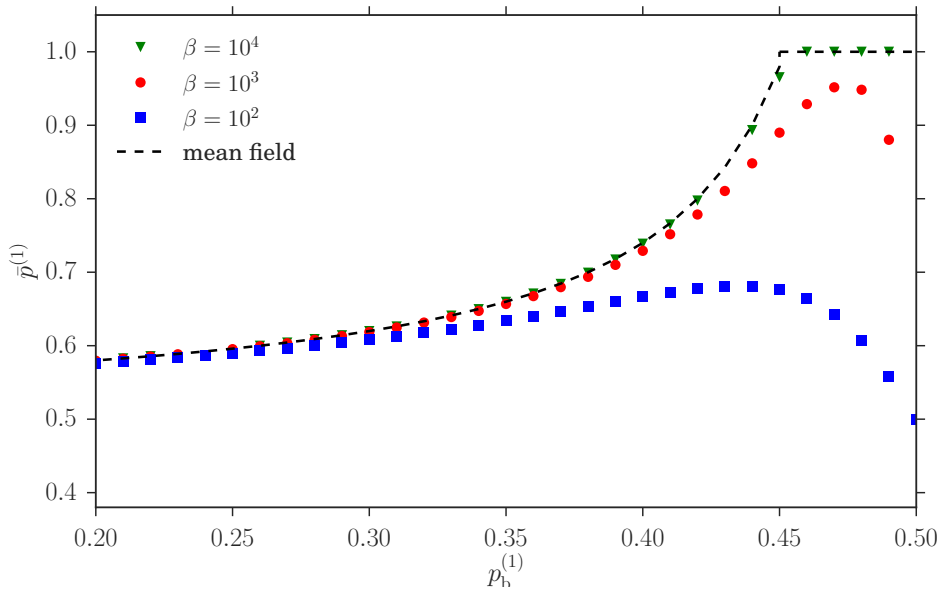

Figure 2.7: Comparison between mean field Nash equilibria (continuous lines) and homogeneous steady states of EWA learning (symbols) for three different values of the intensity of choice β. The market biases are $\theta_1 = 1 - \theta_2 = 0.3$ and the buying probabilities $p_b^{(1)} = 1 - p_b^{(2)} = p_b$. Shown is $\bar{p}^{(1)}$, the fraction of traders from the first class going to the first market, versus $p_b^{(1)}$.

first market in the steady state of EWA learning. The results for three different β are compared to the values of $\bar{p}^{(1)}$ determined from the mean field Nash equilibrium condition, which as we saw leads to the two payoff equalities (2.8). As expected, as β gets larger, the aggregate $\bar{p}^{(1)}$ gets closer to its Nash equilibrium value, confirming our reasoning above. Note around $p_b^{(1)} = 0.45$ we transition from the situation in Fig. 2.2(a), where the Nash equilibrium and the corresponding steady state are of homogeneous mixed type (green triangle in the figure), to the homogeneous pure state (orange square) in Fig. 2.2(b).

So far our main conclusion is that steady states of EWA learning can give *homogeneous mixed* realizations of the potentially heterogeneous Nash equilibria we had identified in Sec. 2.3: even though the equilibrium could be heterogeneous, the dynamics generates a homogeneous steady state with the same

aggregates where all players use the same mixed strategy. This happens if we consider the limit of the dynamics for $\beta \to \infty$ at $\alpha = 0$. One would expect from the phase diagram in Fig. 2.5 that the same steady state is obtained if we move the path of approach towards $(\alpha, 1/\beta) = (0, 0)$ slightly away from the vertical axis, *i.e.* if α is nonzero but goes to zero sufficiently fast as β grows. We show in Appendix 2.D that this is true if the decay of α is exponential, $\alpha_c \sim \exp(-\text{const} \cdot \beta)$: if the constant in the exponent is large enough, the attraction distributions remain homogeneous and attractions again become equal to payoffs for $\beta \to \infty$.

2.4.5 Heterogeneous attraction distributions

We investigate in this section steady states of EWA learning where the attraction distributions of traders are multimodal (heterogeneous) rather than unimodal. As explained in Sec. 2.4.3, for $r \to 0$ the modes become sharp peaks so that unimodal distributions become homogeneous. We have investigated the latter case so far, but heterogeneous steady states should also exist. Indeed, it was shown in [2] using multi-agent simulations as well as theoretical studies of the Kramers-Moyal expansion detailed in Sec. 2.4.2 that for high enough intensity of choice β the distribution of attractions undergoes a transition from homogeneous to heterogeneous. We therefore expect to find heterogeneous steady states of EWA learning more generally for large β and α not too small. We confirm this expectation in this section, where we also find surprising transitions between different types of heterogeneous steady states.

Difference between the case of homogeneous and heterogeneous attraction distributions

In [2], Alorić *et al.* describe a method to obtain the critical α at which the attraction distributions of the traders in the two classes become heterogeneous. One assumes initially that the distributions are homogeneous and determines a self-consistent assignment of the aggregates $\bar{p}^{(1)}$, $\bar{p}^{(2)}$ on this basis. One then checks whether the single agent dynamics for these aggregate values has one fixed point, producing a homogeneous distribution of attractions, or two or more (stable) fixed points, giving a heterogeneous distribution with peaks at these locations in attraction space. What this method leaves open, however, is what the weights of these peaks are and in particular whether they remain nonzero in the large memory limit $r \to 0$. This is the task we tackle using large deviation methods, as summarised in Sec. 2.4.3 above and described in more detail in 2.5.

Transition from one to two to three stable fixed points

We next explore the different fixed point structures of the single agent dynamics as a function of the fictitious play parameter α, for fixed large intensity of choice β. In principle at each α the aggregates $\bar{p}_1^{(c)}$, $\bar{p}_2^{(c)}$ need to be determined from self-consistency but from the experience with the homogeneous solutions we expect that as long as α is small enough and β large enough, the self-consistent aggregate values will be close to their Nash equilibrium values. To leading order one can therefore think of varying α at fixed aggregates. As before we also rely on the assumption that the memory of the traders is large ($r \to 0$); the finite memory case will be investigated below using numerical simulations.

When the fictitious play coefficient α is small enough, the single agent dynamics

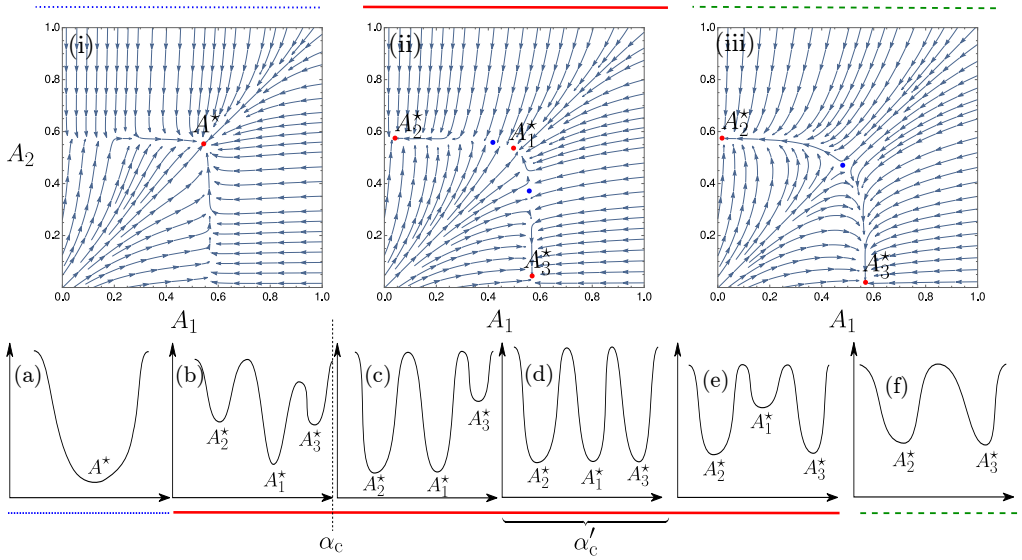

Figure 2.8: (i-iii) Flow diagrams of the single agent dynamics for increasing α. The points represent the stable (red) and unstable (blue) fixed points of the dynamics. The potentials in the bottom row represent schematically in 1-D the arrangement of fixed points (stable = potential minimum, unstable = potential maximum). Attraction distributions are peaked around stable fixed points; in the 1-D representation, the lowest minima indicate peaks with weights of order unity as $r \to 0$, while higher-lying (metastable) minima correspond to peaks that become exponentially suppressed. For $\alpha < \alpha_c$, the aggregates of the single agent dynamics are deduced by self consistency from the only stable fixed point of the dynamics (panels (a) and (b)), while for larger α the aggregates are chosen such that the transition rates between the stable fixed points ($\boldsymbol{A}_1^\star$ and $\boldsymbol{A}_2^\star$ for $\alpha_c < \alpha < \alpha_c'$, panels (c) and (d); $\boldsymbol{A}_1^\star$ and $\boldsymbol{A}_3^\star$ for $\alpha > \alpha_c'$, panels (e) and (f)) are of the same order. Plots were produced with symmetric market biases $\theta_1 = 1 - \theta_2 = 0.3$ and probability of buying $p_b^{(1)} = 1 - p_b^{(2)} = 0.2$ and intensity of choice $\beta = 1/0.11$.

has a single stable fixed point $\boldsymbol{A}_1^\star$ (see Appendix 2.D and Fig. 2.8(i)) and so for $r \to 0$ the distribution of attractions is a δ-peak at this point as shown in Fig. 2.3(a). As α increases then as shown in Fig. 2.8(b) two new stable fixed

points $\boldsymbol{A}_2^\star$ and $\boldsymbol{A}_3^\star$ appear, first one and then the other. But the distribution of attractions is still delta peaked around the original fixed point because in the limit $r \to 0$ the other fixed points are exponentially suppressed in $1/r$: they are in this sense metastable.

The first phase transition arises at a critical value of α, α_c, where one of the metastable point becomes stable; in Fig. 2.8 this is $\boldsymbol{A}_2^\star$. In this case, the attraction distribution is composed of two δ-peaks located at these two stable fixed points of the single agent dynamics (see Fig. 2.3(a,b) for an example projected onto one direction in attraction space). The transition occurs because the actions (see Sec. 2.5) for single agents to move from one stable fixed point to the other and for the reverse move become equal.

This ensures that the fluxes of agents between the two stable fixed points are of the same order of magnitude in both directions, and hence that the two peaks in the attraction distribution can have comparable rather than exponentially different weights.

As α increases further, small changes to the aggregates maintain the condition of comparable flux between the two existing stable peaks. Eventually, at some α_{c}' higher than α_c, the third fixed point also becomes stable so that the attraction distribution acquires three peaks.

Note that the weights of the three peaks cannot be fully determined at $\alpha = \alpha_{\mathrm{c}}'$: the self-consistency for $\bar{p}^{(1)}$ only gives one condition for three nonnegative peak weights that need to sum to one, so that the problem is underconstrained. This indicates that for nonzero r these weights would vary continuously across a small range of α of order r.

For $\alpha > \alpha_{\mathrm{c}}'$, it is the turn of the central fixed point $\boldsymbol{A}_1^\star$ to become metastable; aggregate values are determined by the equal action condition between the two outer stable fixed points and the attraction distribution goes back to having

only two δ-peaks. Finally at even larger α the central metastable fixed point disappears altogether in a saddle-node bifurcation.

Game theoretical interpretation of the steady states

We now investigate the characteristics of all the steady states described above and compare each of them to the Nash equilibria enumerated in Sec. 2.3. When α is below the critical value α_c, all the traders within one class randomise between the two markets, going to the first market with the same probability. This probability is $\sigma_\beta(A_1^{(c)} - A_2^{(c)})$ evaluated at the stable fixed points of the single agent deterministic dynamics, which also equals $\bar{p}^{(c)}$ (see Sec. 2.4.1). This *homogeneous mixed* strategy profile is plotted as the single-peaked preference distribution in Fig. 2.3.

For the opposite case of large α, $\alpha > \alpha_c'$, there are within each class two sub-populations of traders, each of which corresponds to a peak of the attraction distribution as shown schematically in Fig. 2.3(b). Looking at Fig. 2.8(iii) and (f), one sees that at both of these peaks, the attractions to the two markets remain distinct for large β – the relevant fixed points are far from the 45° diagonal. In the limit both sub-populations will therefore play a pure strategy as $\sigma_\beta(A_1^{(c)} - A_2^{(c)})$ tends to one or zero, respectively. This situation is shown as the preference distribution in Fig. 2.3(b) with two peaks around preference one and zero, representing two sub-populations of traders all choosing market 1 and 2 respectively. This steady state of EWA learning is therefore a *heterogeneous pure* realization of a Nash equilibrium, as the preferences of traders are heterogeneous, with two sub-population playing different pure strategies.

While the two cases of homogeneous mixed and heterogeneous pure Nash equilibria are well studied in the literature [90, 15], we find a novel state for $\alpha_c < \alpha < \alpha_c{}'$. Again there are within each class two sub-populations of

traders. But now one sub-population has attractions that become equal for large β: the corresponding fixed point lies close to the diagonal in Fig. 2.8(ii). These traders therefore play a mixed strategy and randomise between the two markets. Overall we have a *heterogeneous mixed* steady state because not all traders play pure strategies. This is illustrated in the right panel of Fig. 2.3. Such heterogeneous mixed strategy distributions have, to our knowledge, never been reported in any study of aggregative games so it is fascinating that they are accessible by EWA learning dynamics.

Overall, we have found that potentially heterogeneous Nash equilibria can be realised as steady states of EWA learning in three different ways by appropriately taking the limits of perfect fictitious play $\alpha \to 0$ and best response $\beta \to \infty$. For small enough $\alpha < \alpha_c(\beta)$ one obtains a homogeneous mixed equilibrium, while keeping larger $\alpha > \alpha_c'(\beta)$ gives a heterogeneous pure equilibrium. Most interesting is the case where α is taken to zero in the "corridor" $\alpha_c < \alpha < \alpha_c'$, which results in a heterogeneous mixed equilibrium.

Note that the partially heterogeneous Nash equilibria (where one class of traders splits into sub-populations while the other stays homogeneous) do not appear in the analysis above because we restricted ourselves to studying Nash equilibria for which the aggregates are symmetric ($\bar{p}^{(1)} = 1 - \bar{p}^{(2)}$), thus ruling out partially heterogeneous Nash equilibria.

We close this section by showing in Fig. 2.9 some numerical results for the aggregate $\bar{p}^{(1)}$ as a function of α, for a fixed intensity of choice β. The values of α_c and α_c' are shown to indicate the transitions between the homogeneous mixed, heterogeneous mixed and heterogeneous mixed states as α grows. Also shown is the even larger critical value α_c'' at which the "central" fixed point (see Fig. 2.8) disappears. Note the vertical scale of the plot, which demonstrates a key point: even though $\beta = 1/0.11$ is not yet very large, $\bar{p}^{(1)}$ is already

quite close to the value $\bar{p}^{(1)} \approx 0.42$ for the potentially heterogeneous Nash equilibrium as calculated using the equal payoff criterion (2.8) in Sec. 2.3.

As we have argued this agreement should get even better as β grows. Numerical data supporting this are shown in Fig. 2.10: $\bar{p}^{(1)}$ decreases towards the Nash equilibrium value with increasing β. Also displayed are the critical values α_c and α_c', which as expected tend to zero as β grows. It is these values that were used to produce the phase diagram in Fig. 2.5.

We note as an aside that in Fig. 2.10 the variation of $\bar{p}^{(1)}$ with α is rather steeper in the heterogeneous mixed phase (between α_c and α_c') than in the homogeneous mixed regime. This probably reflects the change in the way the aggregates are determined in the two regimes: in the homogeneous-mixed phase the aggregates are obtained only by the self-consistency condition for the fixed point location, while they are fixed by the equal flux condition in the heterogeneous mixed phase.

Test against simulations

In this section we test the theoretical predictions obtained above in the $r \to 0$ and for infinite population size N against agent based simulations with a finite memory ($r > 0$) and finite N. We are primarily interested in the steady state of the attraction distribution of the agents, but also consider its time evolution to this steady state. We continue to consider symmetric scenarios so focus on the properties of agents of class 1 throughout. Depending on where the key parameters α and β are in the phase diagram of Fig. 2.5, one expects qualitatively different shapes for the attraction distribution resulting from the learning dynamics. We present simulation results in each of the distinct regions of the phase diagram in Fig. 2.5.

The first zone of interest is on the far left of the phase diagram, where α

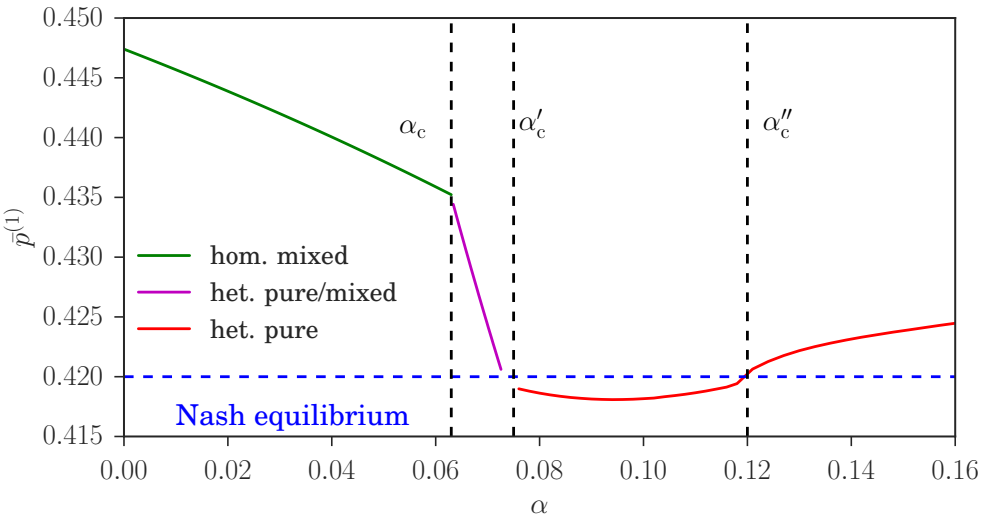

Figure 2.9: Fraction of traders from the first class in the first market, $\bar{p}_1^{(1)}$, for intensity of choice $\beta = 1/0.11$, compared with the value of $\bar{p}^{(1)}$ calculated for the corresponding potentially heterogeneous Nash equilibrium (see Sec. 2.3). Note that the deviation between the two values is small throughout. Critical values of α separating the different types of steady states are indicated; α''_c is the value of α where the "central" fixed point representing traders playing mixed strategies disappears. Same system parameters as in Fig. 2.8.

is below the first segregation threshold α_c. Here, in the steady state of the learning dynamics, we observe in Fig. 2.11(c) the homogeneous distribution of preference predicted by the theory. Looking beyond this agreement for the steady state at the time evolution, panel 2.11(a) shows that for $r = 0.005$ the transient dynamics of the aggregates is nonetheless different from the homogeneous population deterministic dynamics. This appears to be related to a transient segregation effect observed in a small time window around $t = 10$ (Fig. 2.11(b)). This transient segregation does not occur for lower values of r (*e.g.* $r = 0.001$), where the dynamics of the aggregates is closer to the homogeneous population dynamics (see Fig. 2.11(a)).

When $\alpha \in [\alpha_c, \alpha'_c]$, the aggregates relax close to their value in a Nash equi-

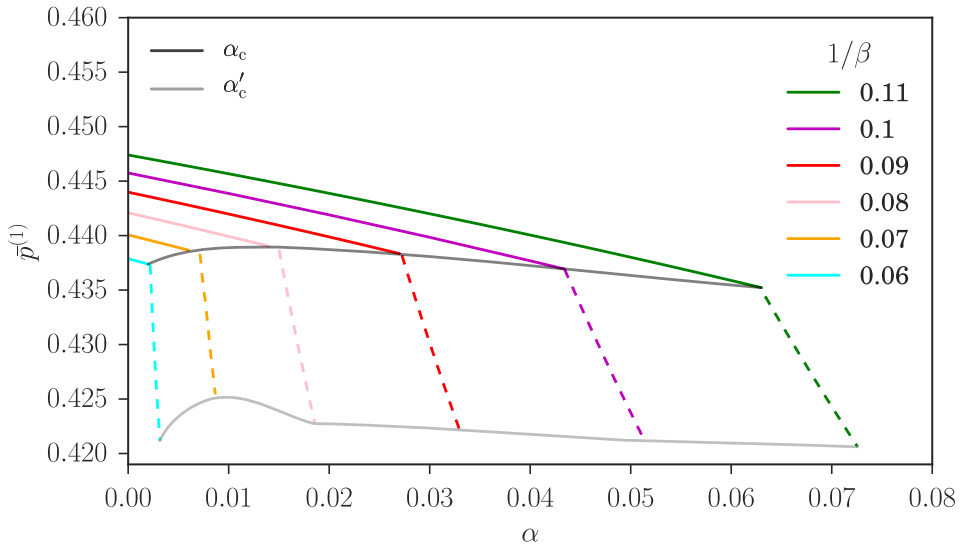

Figure 2.10: Fraction of players from class 1 in the first market, $\bar{p}^{(1)}$, for different values of β. The grey lines connect the values at the two critical α (see Fig. 2.8) as a guide to the eye. System parameters as in Fig. 2.8. Note that $\bar{p}^{(1)}$ gets progressively closer to the Nash equilibrium value ≈ 0.42 as the intensity of choice β grows.

librium around which they fluctuate. Then, they escape from this state to reach an heterogeneous pure Nash equilibrium. The time they remain close to the Nash equilibria depends on the number of agents in the simulation as shown in Fig. 2.13. The theory predicts a distribution composed of two peaks, one peak corresponding to a sub-population playing mixed strategies and the second one to a sub-population playing pure strategies. The results of our simulation presented in Fig. 2.12(a) show a preference distributions composed of three peaks, not two as the theory predicts. On also notices that while the theoretical predictions for the location of the peaks are consistent with the simulation results, the width of the peaks in the simulations is larger than predicted. We believe this is because the theoretical predictions for the width of the peaks make the assumption that the system is in its steady state. This is not strictly verified here as the finite-N system is in a transient state before relaxing to a heterogeneous pure distribution of strategies. As α goes

above the second segregation threshold, α_c', the dynamics initially continues to show three peaks, but in qualitative agreement with the theory the size of the central peak diminishes rapidly, becoming negligible for large enough α. The preference distributions obtained from simulations are then consistent with the theoretical predictions as shown in Fig 2.12(d). Moreover, the aggregates stay close to $f_1 = 0.42$ and never diverge to $f_1 = 1$ or $f_1 = 0$ (as happens for lower values of α).

In summary, the simulations are in good qualitative accord with the predicted sequence of steady states for increasing α: homogeneous mixed, heterogeneous mixed (outer and central peak), heterogeneous mixed (three-peaked) and finally heterogeneous pure (two outer peaks). Corrections to the theoretical predictions arise from the fact that some steady states have a lifetime that only becomes infinite for $N \to \infty$, and from the use of nonzero r in the simulations.

2.5 Large deviation methods

We describe in this section the large deviation methods we use to study heterogeneous attraction distributions in the steady state of EWA learning. As explained in Sec. 2.4, steady state attraction distributions for small r will be peaked around the stable fixed points of the single agent dynamics. The shape of these peaks becomes Gaussian for $r \to 0$, with a covariance matrix proportional to r that is straightforward to determine. Much more difficult to find are the *weights* of the peaks as these involve rare fluctuations of an agent making the transition from one peak to another. In one dimension the problem is tractable as an explicit formula for the steady state distribution of attractions can be given [2]. In higher dimensions detailed balance [80] has a similar sim-

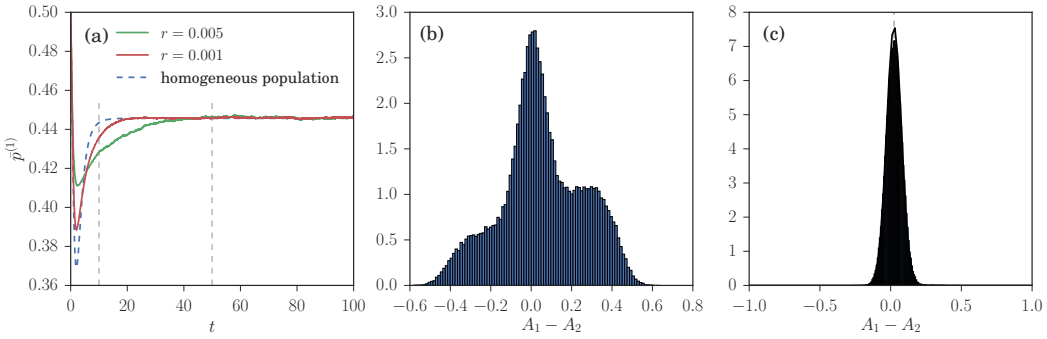

Figure 2.11: EWA learning dynamics at small $\alpha = 0.01$. (a) Time evolution of $\bar{p}^{(1)}$ for $r = 0.005$ and $r = 0.001$ compared to the homogeneous population dynamics predicted for ($r \to 0$). (b,c) Distribution of attraction differences across traders of class 1 at two times, for $r = 0.005$. Black lines are theoretical predictions based on the homogeneous population dynamics and agree well at small r and late times t as expected (see text). Note that for the larger r, the dynamics (a) and the attraction distributions (b) deviate from the small-r theory, showing a transient segregation behaviour that is the precursor of steady state segregation (see Fig. 2.12 (d)) at larger α. The parameters used for those simulation are $\beta = 1/0.11$, $\theta_1 = 1 - \theta_2 = 0.3$, $p_{\mathrm{b}}^{(1)} = 1 - p_{\mathrm{b}}^{(2)} = 0.2$, the system is composed of 20000 traders.

plifying effect, but our single agent dynamics in the two-dimensional attraction space (for each class of agents) does not have this property.

In our approach we consider the peak weights in an attraction distribution as a result of the balance between transitions between the various peaks. We therefore need to find the rates for these transitions. To do this, note from the Kramers-Moyal expansion sketched in Sec. 2.4.2 that the single agent EWA learning is described by a Langevin equation with noise variance $O(r)$. For $r \to 0$ we are therefore looking for transition rates in a low noise limit. This allows us to use Freidlin-Wentzell theory, which deals the with large deviations of Langevin dynamics in exactly this limit [38].

Freidlin-Wentzell theory

We use Freidlin-Wentzell theory in the form developed in [11, 12], which generalises, the Eyring-Kramers [58] formula for the rates of noise-activated transitions to non-conservative dynamics such as our EWA learning. We give a brief summary of those aspects of Freidlin-Wentzell theory that we use in our numerical application and refer to [38] for a mathematically rigorous description and to [11] for a more statistical physics-oriented summary.

Freidlin-Wentzell theory is concerned with the transition rates between two stable states (here $\boldsymbol{A}_1^\star$ and $\boldsymbol{A}_2^\star$) of a non-conservative stochastic dynamics in the low noise limit. A general Langevin equation can be written in the form

$$\dot{\boldsymbol{A}}^{(c)}(t) = \boldsymbol{\mu}^{(c)}(\mathbf{A}^{(c)}(t), \bar{p}^{(1)}, \bar{p}^{(2)}) + \sqrt{r}\boldsymbol{\Sigma}^{(c)\,1/2}(\mathbf{A}^{(c)}(t), \bar{p}^{(1)}, \bar{p}^{(2)})\boldsymbol{\xi}(t) \quad (2.20)$$

where $\boldsymbol{\xi}(t)$ is white noise with unit covariance matrix. The drift $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ of the noise in the Langevin equation are given in Appendix 2.C for our specific case of EWA learning, where the Langevin description results from a second order Kramers-Moyal expansion (2.44). In the generic version above we have omitted the superscript (c) indicating the class of agents we are considering, as well as the dependence of drift and noise covariance on the aggregates $\bar{p}^{(1)}$ and $\bar{p}^{(2)}$.

Associated with the Langevin dynamics is an Onsager-Machlup action $\mathcal{S}[\boldsymbol{A}]$ for any path $\boldsymbol{A}(t)$:

$$\mathcal{S}[\boldsymbol{A}] = \int_{t_1}^{t_2} \frac{1}{2}\left(\dot{\mathbf{A}}(t) - \boldsymbol{\mu}(\mathbf{A}(t))\right)^T \boldsymbol{\Sigma}^{-1}(\mathbf{A}(t))\left(\dot{\mathbf{A}}(t) - \boldsymbol{\mu}(\mathbf{A}(t))\right) \mathrm{d}t \quad (2.21)$$

The action determines the probability of observing any path $[\boldsymbol{A}(t)]$ according to

$$\Gamma_{1\to 2} \sim \exp(-\mathcal{S}[\mathbf{A}]/r) \quad (2.22)$$

where $\sim$ means that the equality is true up to a pre-factor (which depends on the time discretisation used).

The main Freidlin-Wentzell result we need is that the rate $\Gamma_{1\to2}$ for a transition from $\boldsymbol{A}_1^\star$ to $\boldsymbol{A}_2^\star$ (*forward path*) is [38, 14]

$$\Gamma_{1\to2} \sim \exp(-\mathcal{S}^\star_{1\to2}/r) \tag{2.23}$$

where $\mathcal{S}^\star_{1\to2}$ is the minimal action achievable by any paths from $\boldsymbol{A}_1^\star$ to $\boldsymbol{A}_2^\star$ in the infinite time interval $(t_1, t_2) = (-\infty, \infty)$. The rate $\Gamma_{2\to1}$ for the *reverse* transition from $\boldsymbol{A}_2^\star$ to $\boldsymbol{A}_1^\star$ is similarly $\Gamma_{2\to1} \sim \exp(-\mathcal{S}^\star_{2\to1}/r)$.

The attraction distributions we are after will consist of narrow (for small r) peaks at $\boldsymbol{A}_1^\star$ and $\boldsymbol{A}_1^\star$. The weights ω_1 and ω_2 of these two peaks, which represent the probability for an agent to be within each peak, must then be such that forward and backward transitions balance:

$$\omega_1\Gamma_{1\to2} = \omega_2\Gamma_{2\to1} \tag{2.24}$$

$$\frac{\omega_1}{\omega_2} \propto \exp\left(\frac{\mathcal{S}^\star_{1\to2} - \mathcal{S}^\star_{2\to1}}{r}\right) \tag{2.25}$$

This expression shows that when the forward and backward minimal actions are not equal, then one of the two peaks will have an exponentially small weight as $r \to 0$. In practice this is true when the action difference inside the exponential in (2.24) is large compared to r. If it is only of order r or smaller, then we cannot say anything about the weights as we do not determine prefactor in (2.24), though we would expect them to be of order unity.

Finding the minimal action path numerically

Following the method of Bunin *et al.* [14], we find the minimal action by discretising the path $[\boldsymbol{A}(t)]$, evaluating the action as a function of this discretised path and then minimising with respect to the (discretised) path. The path is

discretised into 10 equally spaced timesteps between $t = 0$ and $t = 10$; we found this choice of parameters to be a reasonable trade-off between the precision of our result and the complexity of minimising the discretised action.

There are other methods for finding the minimal value of the action defined in Eq. (2.21), such as solving a Hamilton-Jacobi equation [11], but we chose to use the path discretisation method because we found this to be more robust with respect to changes of model parameters. The discretisation approach could also be improved further, using for example the geometric minimum action method [48], but we found that this was not necessary to achieve the desired precision. We tested this e.g. by benchmarking against closed-form results that can be obtained for $\alpha = 1$ [2].

The numerical path optimisation can be simplified by restricting attention to the *activation* part of the path. Generally, for a system with two stable fixed points $\boldsymbol{A}_1^\star$ and $\boldsymbol{A}_2^\star$ and one saddle point $\bar{\boldsymbol{A}}$ between them, the optimal path starting from $\boldsymbol{A}_1^\star$ will pass through the saddle point $\bar{\boldsymbol{A}}$ and then relax to $\boldsymbol{A}_2^\star$ following the relaxation dynamics $\dot{\boldsymbol{A}}(t) = \boldsymbol{\mu}(\boldsymbol{A}(t))$, as sketched in Fig. 2.14 [38]. Eq. (2.21) shows that the relaxation dynamics does not contribute to the total action as the integrand (the Lagrangian) vanishes identically along this section of the path. As a consequence, the problem of finding a minimal action path between $\boldsymbol{A}_1^\star$ and $\boldsymbol{A}_2^\star$ can be reduced to finding the minimal action path between $\boldsymbol{A}_1^\star$ and $\bar{\boldsymbol{A}}$, *i.e.* from the initial fixed point to the saddle. This restriction significantly improves the precision of the numerical path optimisation.

With the above method, we can work out the action difference between any two fixed points of the single agent dynamics, as a function of the aggregates $\bar{p}^{(1)}$, $\bar{p}^{(2)}$; only the first of these is needed for symmetric steady states. The values of $\bar{p}^{(1)}$ where the action difference between two single agent fixed points vanishes identify the points where the steady state attraction distribution of EWA learning can have more than one peak. Either side of these values,

a single peak is dominant in the attraction distribution; which peak this is changes discontinuously at a zero action difference value of $\bar{p}^{(1)}$, see Fig. 2.6.

2.6 Conclusion

In this chapter we studied a minimal model of agents choosing between two double auction markets, which is a special case of a large aggregative game. Previous work studying EWA (experience weighted attraction) learning in this system had found segregation, where a group of identical agents becomes heterogeneous by separating into sub-groups adopting different behaviours. We first asked the question of whether this phenomenon has an analogue in the Nash equilibria of the corresponding game, where – in contrast to the EWA dynamics – agents have full information about their expected payoffs.

In a game theoretical analysis (Sec. 2.3) we addressed this question within a setup where there are two classes of agents that typically buy and sell, respectively. We showed that two *aggregate* quantities, namely, the fraction of agents from each class choosing the first market, are sufficient to assess whether a distribution of strategies, *i.e.* market preferences, across the agents in each class is a Nash equilibrium or not. This allowed us to classify the Nash equilibria, according to the type of strategies played by the agents (pure or mixed) and according to the distribution of strategies being homogeneous (the entire class population plays the same strategy) or heterogeneous (the population is divided into subpopulations playing different strategies). The model parameters for which each of these Nash equilibria exists are summarised in Fig. 2.4. A key conclusion is that there are regions of heterogeneous equilibria: these are the equilibrium analogues of dynamical segregation as observed previously.

This answer to our first question had to be qualified, however, because there is in general an infinity of strategy distributions consistent with a given pair

of aggregate values. The Nash equilibrium analysis can therefore only identify equilibria as *potentially* heterogeneous but leaves open the nature of the actual strategy distribution, which could be homogeneous mixed, heterogeneous pure or heterogeneous mixed. We therefore asked a second question of whether EWA learning can resolve this ambiguity, by identifying which Nash equilibria can be reached dynamically. We first argued that steady states of EWA learning should be Nash equilibria in the limit of perfect fictitious play ($\alpha \to 0$), long agent memory ($r \to 0$) and best response ($\beta \to \infty$). (Sec. 2.4.1). Non-trivially, however, this joint limit can be taken in several ways, as shown in the phase diagram in Fig. 2.5: depending on how the point $(\alpha, 1/\beta) = (0, 0)$ is approached, a small number of different limiting steady states of EWA learning can result as sketched in Fig. 2.3. These include a homogeneous mixed state, where all agents within a class randomise between markets in the same way, and a heterogeneous pure equilibrium, where agents separate into two groups, each choosing a market deterministically. Along with these standard types of Nash equilibria, however, we also found a *heterogeneous mixed* steady state, where the agents do split into groups but not all groups play deterministically. In fact, at the boundary between the latter two types of steady states (denoted $\alpha = \alpha_c'$ in our analysis) it is possible to generate equilibria where *three* groups of agents appear within each class.

Technically what made our theoretical analysis of the heterogeneous steady states possible was the use of Freidlin-Wentzell theory, which is the tool of choice for studying the behavior of dynamical systems subject to weak noise, here arising from the limit $r \to 0$. We also compared the theoretical results to multi-agent simulations for $r > 0$, finding good qualitative agreement.

While we focused our analysis on the study of the minimal model of choice between double auction market presented in Sec. 2.2, our methods could be applied fruitfully also to the study of EWA learning in other types of aggregative games such as the Cournot model [33]. It would be particularly interesting

to see whether also here dynamical considerations single out particular Nash equilibria, including ones with the novel heterogeneous mixed character that we found in our system.

At a technical level, future work could look more closely at the limit of large intensity of choice β required to realise Nash equilibria as dynamical steady states. We approached this limit numerically, finding good agreement with theoretical predictions already for relatively modest β. An interesting challenge would be to take the full $\beta \to \infty$ limit in closed form within the analysis: preliminary work suggests that the large deviation analysis then becomes rather intricate, hence we leave this aspect for future work.

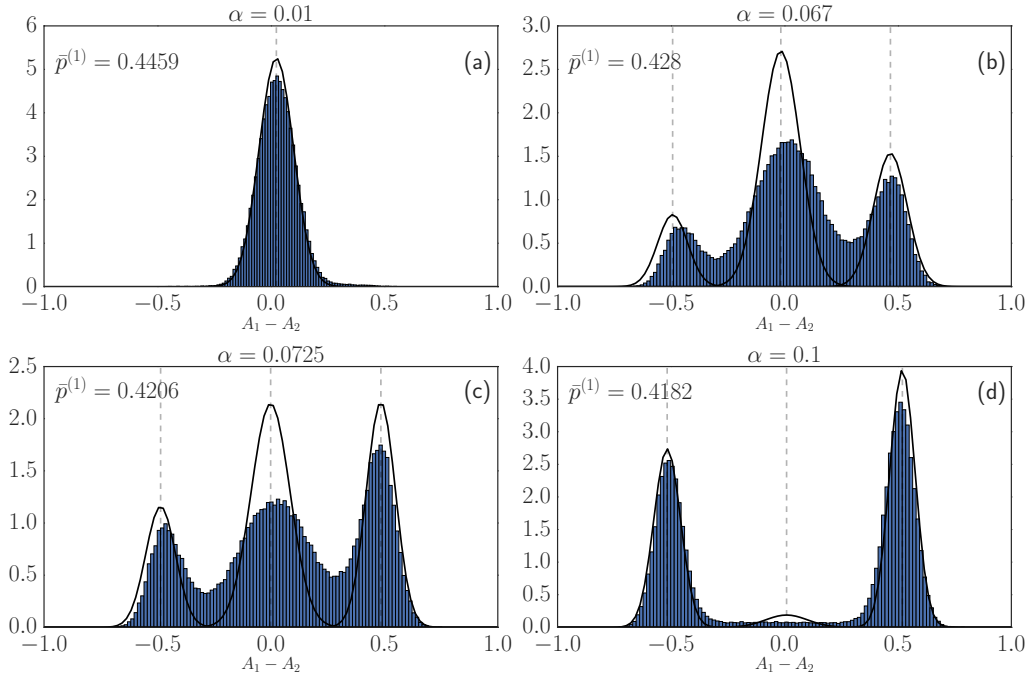

Figure 2.12: Steady state distribution of the attraction differences for $r = 0.01$, and increasing values of α; the remaining parameters are as in Fig. 2.8. When $\alpha = 0.067$ (panel (b)), the theory predicts one outer peak on the right and one inner peak corresponding to a fraction of the population playing a mixed strategy. The simulations additionally show an outer peak on the left, which arises from the fact that the finite-N system is not in a true steady state. Panel (c) shows the situation for $\alpha = 0.0725$, which is the critical value α_c' at which we expect to see from theory three different peaks in the distribution of attraction differences. The theoretical predictions (black curves) is a Gaussian mixture composed of three peaks whose mean and variance are obtained from the Kramers-Moyal expansion while their weights, which the theory cannot predict, are fitted to the data. The peak positions are in good agreement with theory while the simulations overestimate the variance of the peaks, again because of transient effects. In panel (d), for $\alpha > \alpha_c'$, there is very good agreement with theory except for a small central peak that for $r \to 0$ is predicted to have weight zero. This is likely to be an effect of the nonzero $r = 0.01$ used in the simulations.

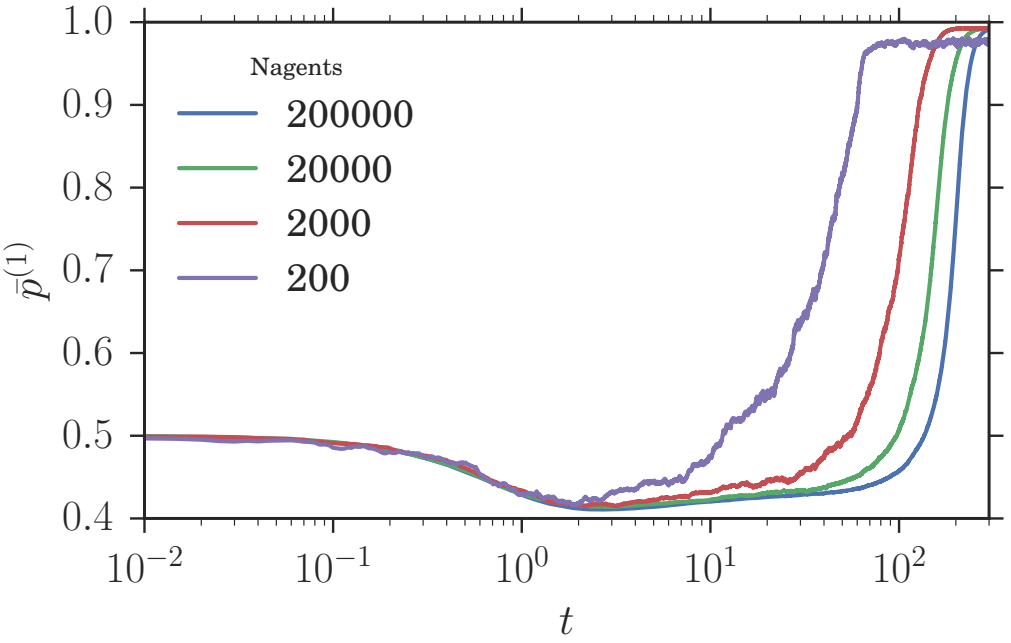

Figure 2.13: Time evolution of $\bar{p}^{(1)}$ for $\alpha = 0.068$, $r = 0.005$ and different numbers of agents N. Other parameters are the same as in Fig. 2.8

.

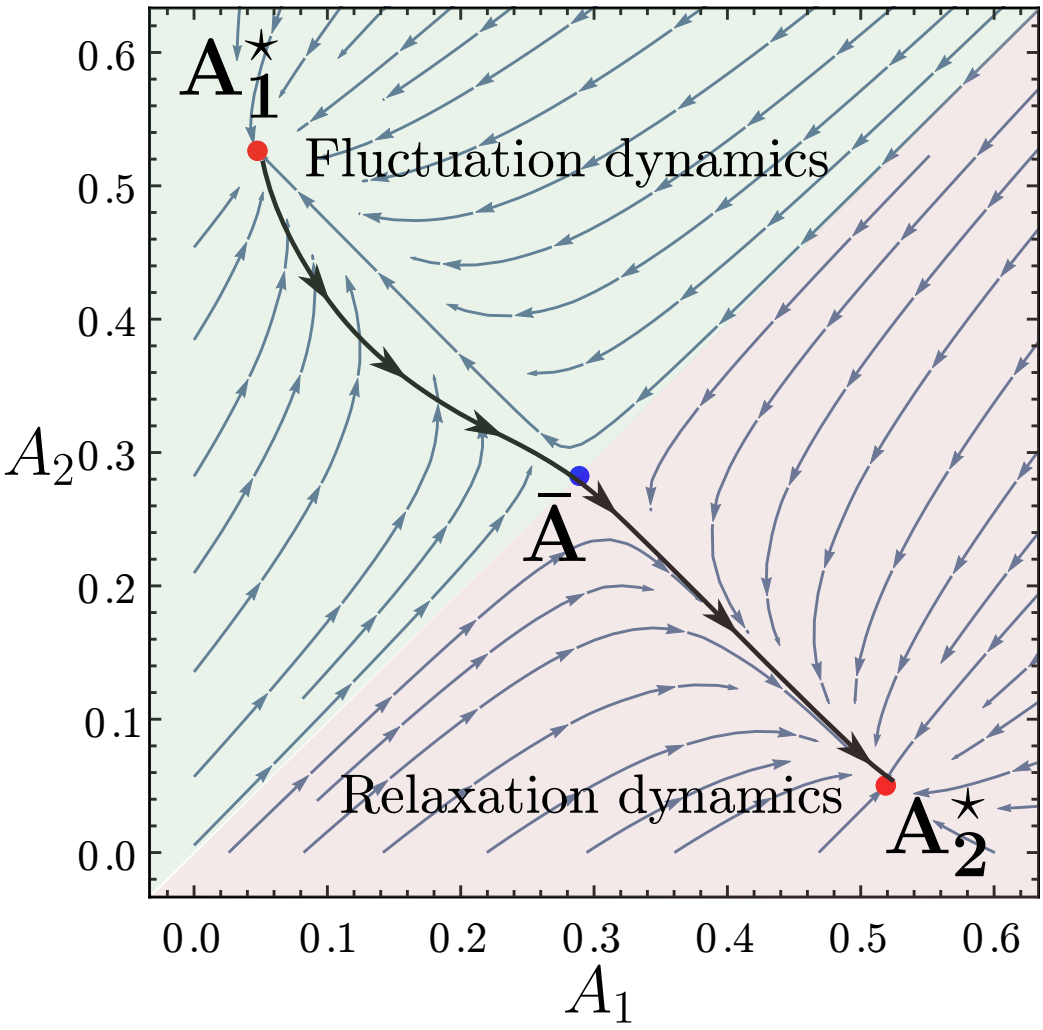

Figure 2.14: An example of a minimal action path, from fixed point $\boldsymbol{A}_1^\star$ to $\boldsymbol{A}_2^\star$. The path starts with a "fluctuation" (or: activation) segment that ends at the saddle point $\bar{\boldsymbol{A}}$ between the two fixed points. The remainder of the path is a "relaxation" segment that follows the deterministic dynamics and incurs zero contribution to the action.

Appendix

2.A Formula for the payoff

To work out the average payoff of an ask (a) or bid (b) at market m, we find first the probability for such an order to be valid:

$$\mathcal{V}(\mathrm{a}, m) = \mathbb{P}(\text{ask price} < \pi_m) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\pi_m} \exp\left(-\frac{(x-\mu_\mathrm{a})^2}{2\sigma^2}\right) \mathrm{d}x \tag{2.26}$$

$$\mathcal{V}(\mathrm{b}, m) = \mathbb{P}(\text{bid price} > \pi_m) = \frac{1}{\sqrt{2\pi}\sigma} \int_{\pi_m}^{\infty} \exp\left(-\frac{(x-\mu_\mathrm{b})^2}{2\sigma^2}\right) \mathrm{d}x \tag{2.27}$$

where the trading price π_m is defined in equation (2.1).

Once an order has been validated, it needs to be matched with that of a trader on the other side of the market. We denote the probability for this to happen for an order of type τ at market m by $\mathcal{M}(\tau, m, f_m)$. This quantity depends on the ratio of the number of buyers and sellers in the market, $f_m = \frac{\#\text{ buyers @ market m}}{\#\text{ sellers @ market m}}$, as follows:

$$\mathcal{M}(\mathrm{a}, m, f_m) = \min\left(\frac{f_m \mathcal{V}(\mathrm{b}, m)}{\mathcal{V}(\mathrm{a}, m)}, 1\right) \tag{2.28}$$

$$\mathcal{M}(\mathrm{b}, m, f_m) = \min\left(\frac{\mathcal{V}(\mathrm{a}, m)}{f_m \mathcal{V}(\mathrm{b}, m)}, 1\right) \tag{2.29}$$

where the first ratio in the minimum is that of the number of *valid* buy and sell orders, always assuming large N where fluctuations of these numbers can be neglected.

We call $\langle \mathcal{S}_{\tau,m} \rangle$ the average score of an order of type τ, once it has been validated and successfully matched. This is given by:

$$\langle \mathcal{S}_{\mathrm{a},m} \rangle = \frac{1}{\mathcal{V}(\mathrm{a},m)} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\pi_m} (\pi_m - x) \exp\left(-\frac{(x-\mu_\mathrm{a})^2}{2\sigma^2} \right) \mathrm{d}x \tag{2.30}$$

$$\langle \mathcal{S}_{\mathrm{b},m} \rangle = \frac{1}{\mathcal{V}(\mathrm{b},m)} \frac{1}{\sqrt{2\pi}\sigma} \int_{\pi_m}^{\infty} (x - \pi_m) \exp\left(-\frac{(x-\mu_\mathrm{b}^2)}{2\sigma^2} \right) \mathrm{d}x \tag{2.31}$$

For later use we also define the average square of the score:

$$\langle \mathcal{S}^2_{\mathrm{a},m} \rangle = \frac{1}{\mathcal{V}(\mathrm{a},m)} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\pi_m} (\pi_m - x)^2 \exp\left(-\frac{(x-\mu_\mathrm{a})^2}{2\sigma^2} \right) \mathrm{d}x \tag{2.32}$$

$$\langle \mathcal{S}^2_{\mathrm{b},m} \rangle = \frac{1}{\mathcal{V}(\mathrm{b},m)} \frac{1}{\sqrt{2\pi}\sigma} \int_{\pi_m}^{\infty} (x - \pi_m)^2 \exp\left(-\frac{(x-\mu_\mathrm{b})^2}{2\sigma^2} \right) \mathrm{d}x \tag{2.33}$$

We can now compute the average payoff of an order of type τ at market m:

$$\mathcal{P}_{\tau,m}(f_m) = \mathcal{V}(\tau,m)\mathcal{M}(\tau,m,f_m)\langle \mathcal{S}_{\tau,m} \rangle \tag{2.34}$$

Similarly, the average squared payoff that will appear in the second order moment of the Kramers-Moyal expansion in App. 2.C can be expressed as

$$\mathcal{Q}_{\tau,m}(f_m) = \mathcal{V}(\tau,m)\mathcal{M}(\tau,m,f_m)\langle \mathcal{S}^2_{\tau,m} \rangle \tag{2.35}$$

$$\mathcal{Q}^{(c)}_{m}(f_m) = p^{(c)}_\mathrm{b}\mathcal{Q}_{\mathrm{b},m}(f_m) + (1 - p^{(c)}_\mathrm{b})\mathcal{Q}_{\mathrm{a},m}(f_m) \tag{2.36}$$

The second version here is averaged over the preference for buying and selling of an agent in class c.

2.B Phase diagram boundaries in Fig 2.4

In this section we indicate how to calculate phase boundaries in Fig. 2.4, which shows the phase diagram for the case where the market bias and the probability to buy are symmetric ($\theta_1 = 1 - \theta_2$, $p_\mathrm{b} \doteq p^{(1)}_\mathrm{b} = 1 - p^{(2)}_\mathrm{b}$).

At this boundary, a (symmetric) potentially heterogeneous Nash equilibrium (green triangle in Fig. 2.2) turns smoothly into a homogeneous pure equilibrium (blue diamond and orange square in Fig. 2.2) where the two classes of

players choose different markets. One can therefore calculate the boundary by establishing the zone in the phase diagram where this homogeneous Nash equilibrium exists. For definiteness we consider the equilibrium $(\bar{p}^{(1)}, \bar{p}^{(2)}) = (1, 0)$; the calculation for $(0, 1)$ is completely analogous.

To get rid of the min in Eq. (2.28, 2.29) we focus in addition on the case where market 1 is saturated with sellers:

$$\frac{f_1 \mathcal{V}(\mathrm{b}, 1)}{\mathcal{V}(\mathrm{a}, 1)} < 1 \tag{2.37}$$

As a consequence the min term disappears from the market conditions:

$$\mathcal{M}(\mathrm{b}, 1, f_m) = \mathcal{M}(\mathrm{a}, 2, f_1) = 1 \tag{2.38}$$

$$\mathcal{M}(\mathrm{a}, 1, f_m) = \mathcal{M}(\mathrm{b}, 2, f_2) = \frac{f_1 \mathcal{V}(\mathrm{b}, 1)}{\mathcal{V}(\mathrm{a}, 1)} \tag{2.39}$$

Here the equality between $\mathcal{M}(\mathrm{a}, 1, f_1)$ and $\mathcal{M}(\mathrm{b}, 2, f_2)$ comes from the symmetry of the parameters. Because $(\bar{p}^{(1)}, \bar{p}^{(2)}) = (1, 0)$, all agents from class 1 go to market 1 and so the buyer-to-seller ratios f_m from (2.4a) are simple to express in terms of p_b:

$$f_1 = \frac{1}{f_2} = \frac{p_\mathrm{b}}{1 - p_\mathrm{b}} \tag{2.40}$$

The payoffs at the two markets for traders from class 1 simplify accordingly:

$$\mathcal{P}_1^{(1)}(f_1) = p_\mathrm{b} \mathcal{V}(\mathrm{b}, 1) \langle S_{\mathrm{b},1} \rangle + (1 - p_\mathrm{b}) \mathcal{V}(\mathrm{a}, 1) \left[\frac{p_\mathrm{b}}{1 - p_\mathrm{b}} \frac{\mathcal{V}(\mathrm{b}, 1)}{\mathcal{V}(\mathrm{a}, 1)} \right] \langle S_{\mathrm{a},1} \rangle \tag{2.41}$$

$$\mathcal{P}_2^{(1)}(f_1) = (1 - p_\mathrm{b}) \mathcal{V}(\mathrm{a}, 2) \langle S_{\mathrm{a},2} \rangle + p_\mathrm{b} \mathcal{V}(\mathrm{b}, 2) \left[\frac{p_\mathrm{b}}{1 - p_\mathrm{b}} \frac{\mathcal{V}(\mathrm{b}, 1)}{\mathcal{V}(\mathrm{a}, 1)} \right] \langle S_{\mathrm{b},2} \rangle \tag{2.42}$$

The factors in brackets are the matching probabilities from (2.39), from which $\mathcal{V}(\mathrm{a}, 1)$ cancels in the first equation and similarly (by symmetry) $\mathcal{V}(\mathrm{a}, 1) = \mathcal{V}(\mathrm{b}, 2)$ in the second.

Our assumed equilibrium $(\bar{p}^{(1)}, \bar{p}^{(2)}) = (1, 0)$ will be a Nash equilibrium if the payoff at market 1 is higher than at market 2 for players from class 1. (By

symmetry, the payoff relation is then reversed for players in class 2.) From the explicit payoff expressions above, this condition can be re-arranged into

$$0 \leq p_{\mathrm{b}}^2 \left(-\langle S_{\mathrm{a},2}\rangle \mathcal{V}(\mathrm{a},2) - \langle S_{\mathrm{b},1}\rangle \mathcal{V}(\mathrm{b},1) - \langle S_{\mathrm{a},1}\rangle \mathcal{V}(\mathrm{b},1) - \langle S_{\mathrm{b},2}\rangle \mathcal{V}(\mathrm{a},2)\right) + p_{\mathrm{b}} \left(\langle S_{\mathrm{b},1}\rangle \mathcal{V}(\mathrm{b},1) + 2\langle S_{\mathrm{a},2}\rangle \mathcal{V}(\mathrm{a},2) + \langle S_{\mathrm{a},1}\rangle \mathcal{V}(\mathrm{b},1)\rangle\right) - \langle S_{\mathrm{a},2}\rangle \mathcal{V}(\mathrm{a},2) \quad (2.43)$$

For given θ_1 all coefficients in this quadratic equation are known so the phase boundaries can be obtained directly as its roots. We plotted these roots in Fig. 2.B.1; note that the boundaries are close to linear but not exactly so. One has to check a posteriori that the assumption (2.37) of market 1 being saturated with sellers is valid, which rules out the bottom "cone" in the figure.

The remainder of the phase diagram in Fig. 2.4 is obtained by the analogous calculation under the assumption that market 1 is saturated with buyers rather than sellers, which yields the bottom "cone" in Fig. 2.B.1 and by finally repeating the overall reasoning for the Nash equilibrium $(\bar{p}^{(1)}, \bar{p}^{(2)}) = (0, 1)$.

2.C Kramers-Moyal expansion

Here we provide the coefficients of the Kramers-Moyal expansion for traders with fixed buy-sell preference, given fictitious play coefficient α and intensity of choice β. The truncation of the Kramers-Moyal expansion at the second order gives the Fokker-Planck equation for the time evolution of the attraction distributions:

$$\partial_t \mathbb{P}(\mathbf{A}^{(c)}, t) = -\sum_{1\leq m\leq 2} \partial_{A_m^{(c)}} [\mu_m^{(c)}(\mathbf{A}^{(c)}, \bar{p}^{(1)}, \bar{p}^{(2)}) \mathbb{P}(\mathbf{A}^{(c)}, t)] + \frac{r}{2} \sum_{1\leq m,m'\leq 2} \partial^2_{A_m^{(c)} A_{m'}^{(c)}} [\Sigma^{(c)}_{mm'}(\mathbf{A}^{(c)}, \bar{p}^{(1)}, \bar{p}^{(2)}) \mathbb{P}(\mathbf{A}^{(c)}, t)] \quad (2.44)$$

To lighten the notation we will in the following drop the superscript (c) indicating the class of an agent and also suppress the dependence on the aggregates $\bar{p}^{(1)}, \bar{p}^{(2)}$, which are in general time-dependent via Eq. (2.15).

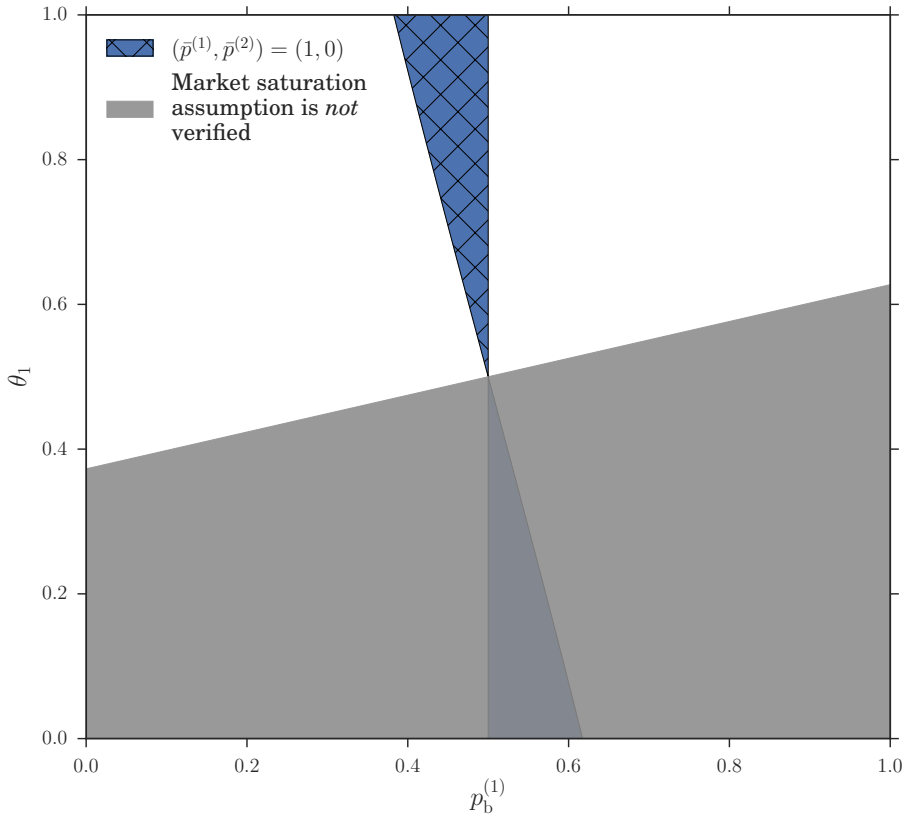

Figure 2.B.1: Analytic determination of boundaries for the zone where a homogeneous Nash equilibrium exists where players from the two classes choose different markets. Within the blue regions the payoff inequality (2.8) is satisfied. The region shaded grey is ruled out by the assumption of market 1 being saturated with sellers.

In the above expansion time has been rescaled as $t = rn$, where n is the number of trading rounds. The time interval $\Delta t = r$ then features in the normalization of the drift and diffusion matrix, which are determined as the first and second order jump moments:

$$\boldsymbol{\mu} = \frac{1}{r}\langle \Delta \mathbf{A} \rangle, \qquad r\boldsymbol{\Sigma} = \frac{1}{r}\langle \Delta \mathbf{A}\, \Delta \mathbf{A}^{\mathrm{T}} \rangle \tag{2.45}$$

where $\Delta \mathbf{A} = \mathbf{A}(n+1) - \mathbf{A}(n)$ is the change in the agent's attraction vector in one training round and the T superscript indicates vector transpose. Writing

$\Delta\mathbf{A}$ explicitly from (2.2) then gives for the drift term:

$$\mu_1(\mathbf{A}) = [\mathcal{P}_1(f_1) - A_1]\sigma_\beta(A_1 - A_2) - \alpha A_1\sigma_\beta(A_2 - A_1) \tag{2.46}$$

$$\mu_2(\mathbf{A}) = [\mathcal{P}_2(f_2) - A_2]\sigma_\beta(A_2 - A_1) - \alpha A_2\sigma_\beta(A_1 - A_2) \tag{2.47}$$

In the diffusion term Σ_{ij} the second order moments of the score distribution also feature, as follows:

$$\begin{aligned}\Sigma_{11}(\mathbf{A}) = &\left[\mathcal{Q}_1(f_1) - 2A_1\mathcal{P}_1(f_1) + {A_1}^2\right]\sigma_\beta(A_1 - A_2)\\ &+ \alpha^2{A_1}^2\sigma_\beta(A_2 - A_1)\end{aligned} \tag{2.48}$$

$$\begin{aligned}\Sigma_{22}(\mathbf{A}) = &\left[\mathcal{Q}_2(f_2) - 2A_2\mathcal{P}_2(f_2) + {A_2}^2\right]\sigma_\beta(A_2 - A_1)\\ &+ \alpha^2{A_2}^2\sigma_\beta(A_1 - A_2)\end{aligned} \tag{2.49}$$

$$\begin{aligned}\Sigma_{12}(\mathbf{A}) = &-\alpha\Big[\mathcal{P}_1(f_1)A_2\sigma_\beta(A_1 - A_2)\\ &+ \mathcal{P}_2(f_2)A_1\sigma_\beta(A_2 - A_1) - A_1A_2\Big]\end{aligned} \tag{2.50}$$

$$\Sigma_{21}(\mathbf{A}) = \Sigma_{12}(\mathbf{A}) \tag{2.51}$$

2.D Fixed points of single agent dynamics

We show here generally that the single agent dynamics can have up to five fixed points, which can be determined from a single nonlinear equation. As before we drop the superscript (c) for the agent class. The aggregates and hence the expected payoffs $\mathcal{P}_1$, $\mathcal{P}_2$ are fixed.

Fixed points are found from the condition that the drift (2.46,2.47) must vanish:

$$0 = (\mathcal{P}_1 - A_1)\sigma_\beta(A_1 - A_2) - \alpha A_1\sigma_\beta(A_2 - A_1) \tag{2.52}$$

$$0 = (\mathcal{P}_2 - A_2)\sigma_\beta(A_2 - A_1) - \alpha A_2\sigma_\beta(A_1 - A_2) \tag{2.53}$$

Writing $\Delta = A_1 - A_2$ and using $\sigma_\beta(A_2 - A_1) = 1 - \sigma_\beta(\Delta)$, one can express A_1 and A_2 in terms of Δ:

$$A_1 = \frac{\mathcal{P}_1 \sigma_\beta(\Delta)}{\sigma_\beta(\Delta) + \alpha[1 - \sigma_\beta(\Delta)]} = \frac{\mathcal{P}_1}{1 + \alpha \exp(-\beta\Delta)} \tag{2.54}$$

$$A_2 = \frac{\mathcal{P}_2[1 - \sigma_\beta(\Delta)]}{1 - \sigma_\beta(\Delta) + \alpha\sigma_\beta(\Delta)} = \frac{\mathcal{P}_2}{1 + \alpha \exp(\beta\Delta)} \tag{2.55}$$

Taking the difference gives a single equation for Δ, which takes a suggestive form if we write $\alpha = \exp(-a\beta)$:

$$\Delta = \frac{\mathcal{P}_1}{1 + \exp(-\beta(\Delta + a))} - \frac{\mathcal{P}_2}{1 + \exp(\beta(\Delta - a))} \tag{2.56}$$

The solutions of this equation, and hence the single agent fixed points, can be obtained graphically by intersecting a straight line (the l.h.s. of Eq. (2.56)) with the function of Δ on the r.h.s. This function has a simple shape as it is the sum of two sigmoids, one increasing from zero to $\mathcal{P}_1$ around $\Delta = -a$ and the other increasing from $-\mathcal{P}_2$ to zero around $\Delta = a$. From the resulting shape, shown in Fig. 2.D.1, at most five intersections with the diagonal can occur.

We are most interested in the limit of large intensity of choice β, where the sigmoids become step functions. For small α, *i.e.* large a, the only solution is then $\Delta = \mathcal{P}_1 - \mathcal{P}_2$. As α is increased and hence a is decreased, the sigmoidal steps move closer to the origin, each creating an additional pair of solutions when a equals the relevant payoff (see Fig. 2.D.1). For large β, one therefore has as transition from one to three (two stable, one unstable) fixed points at

$$\alpha \sim \exp(-\max(\mathcal{P}_1, \mathcal{P}_2)\beta) \tag{2.57}$$

and from three to five (three stable, two unstable) fixed points at

$$\alpha \sim \exp(-\min(\mathcal{P}_1, \mathcal{P}_2)\beta) \tag{2.58}$$

At finite β the fixed points are shifted away from $\Delta = \pm a$ and this would give corrections to a of order $1/\beta$, which would in turn determine the prefactors of

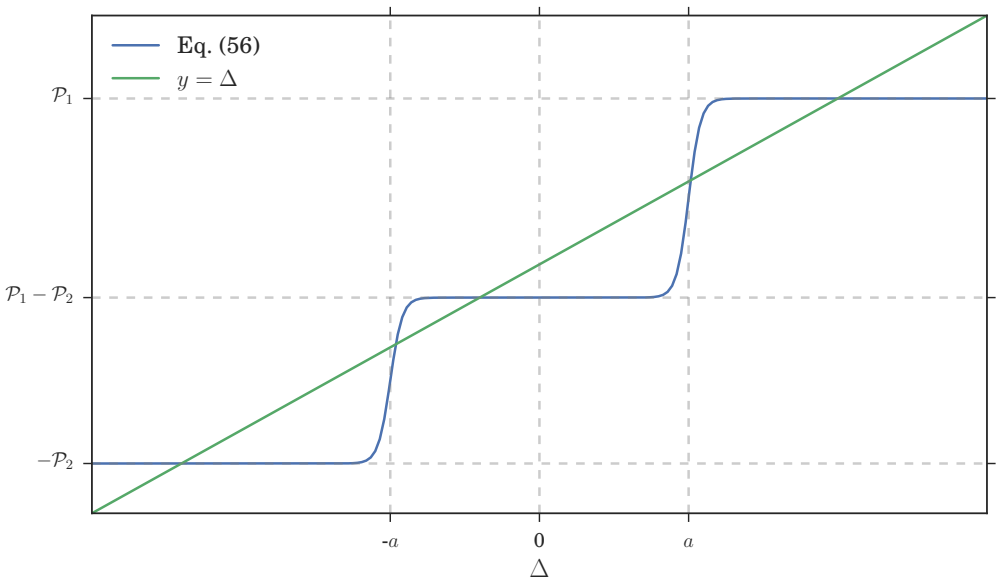

Figure 2.D.1: Sketch of the right handside of the fixed point equation (2.56) for

Δ

the above scalings. Note that as a decreases further, the two sigmoidal ramps will eventually overlap when a is of order $1/\beta$, signalling a transition back to three (two stable) fixed points.

We show in Fig. 2.5 that the scaling of the above α-values, taken at equal payoffs $\mathcal{P}_1 = \mathcal{P}_2$ as is relevant for Nash equilibria, also gives a good account of the variation with β of α_c and α_c'. This suggests that the α-values where new fixed points appear, and where they contribute as peaks with weights of order unity to the steady state distribution, are relatively close, maybe only within a constant prefactor of each other.

CHAPTER 3

Segregation across three markets

3.1 Introduction

In the previous chapter, we studied the effect of using fictitious play on the phenomenon of segregation. To keep this study as simple as possible, we limited it to two double auction markets. However, in general (see the introduction of [91]) traders have more than one market at their disposal and this feature was also present in the CAT game [16] where more than two markets were competing to attract traders. This phenomenology is our motivation to extend the double auction markets model from two to three markets.

A natural approach to studying the coexistence of many markets is multi-agent simulation. Indeed, there is a large body of work that uses the *JCAT* library [72] to explore competition between *continuous double auction markets* [17, 73, 65]. In a spirit similar to our work they use simple learning algorithms such as Zero-intelligence [44] or Zero-Intelligence-Plus [27], for both markets and traders and analyse the allocative efficiency of double auction markets when they are competing against each other. One advantage of purely multi-agent based simulations over our methods is that additional layers of

complexity such as *adaptive markets* and *heterogeneous agents* can be added straightforwardly to the model; in contrast, extending the Fokker-Planck modelling equation to more complex models requires a non-trivial amount of work. We would also like to emphasise that while the market mechanisms implemented in the JCAT library are *continuous* double auctions, ours is more similar to a *clearing house* where the clearing process takes place at discrete time steps.

To finish this short and non-exhaustive review of existing work on multiple market coexistence, we cite the work of Alorić [1] where the *weak segregation* of traders across three double auction markets is examined. However, the absence of the minimal action framework that we implemented in the previous chapter did not allow the study of strong segregation in the large memory limit, and only a limited range of parameter settings was explored. This chapter aims to fill this gap by investigating the existence or otherwise of strong segregation. This is particularly interesting as strong segregation is by definition the only type of segregation that persists in the large memory limit.

In Sec. 3.2 we define the extension of the double actions model of chapter 2 to the case where traders learn to choose among three markets. In Sec. 3.3 we study the appearance of segregation among three fair markets. In Sec. 3.4 we then explore several more general parameter settings: (i) two symmetric markets with varying bias and one fair market, (ii) two symmetric markets and a third market with varying bias, (iii) one market biased toward buyers, one fair market and one other market with generic bias. We conclude in Sec. 3.5 and suggest possible future research directions.

3.2 Preliminaries

3.2.1 Learning rules

In this section, we describe our extension of the model described in Chapter 2 where the traders could choose between two markets to a model where traders can choose between three markets. Here the mechanism of each market remains the same as the one described in chapter 2. As in chapter 2 we consider a population of N traders divided into two classes $c \in \{1, 2\}$ of equal size. Each trader learns at which market he prefers to trade by using Experience-Weighted Attraction learning, and buys with fixed probability $p_{\mathrm{b}}^{(c)}$ depending on the class he belongs to.

Each trader has one attraction for each market (3 in total) which he updates at each trading step with the update rules below:

$$A_m(n{+}1) = \begin{cases} (1-r)A_m(n) + r\mathcal{S}(n) & \text{if the agent chose market } m \text{ in round } n \\ (1-\alpha r)A_m(n) & \text{otherwise} \end{cases} \tag{3.1}$$

The equation above is exactly Eq. (2.2) and we choose to restrict our analysis to the case $\alpha = 1$ for reasons we will explain later. Once each preference is updated, traders use the *multinomial logit function* to choose their next action:

$$\mathbb{P}(M = m) = \frac{\exp(\beta A_m)}{\sum_i \exp(\beta A_i)} \tag{3.2}$$

This is a generalised version of the logit function defined in Eq. (2.3) when one needs to choose between more than two options. In the rest of the chapter we will write the probabilities as a function of the two score differences $\Delta A_2 =$

$A_1 - A_2$ and $\Delta A_3 = A_1 - A_3$:

$$\begin{aligned}
\mathbb{P}(M=1) &= \frac{1}{1+\exp(-\beta\Delta A_2)+\exp(-\beta\Delta A_3)} \\
\mathbb{P}(M=2) &= \frac{1}{1+\exp(\beta\Delta A_2)+\exp(-\beta\Delta A_3+\beta\Delta A_2)} \\
\mathbb{P}(M=3) &= \frac{1}{1+\exp(\beta\Delta A_3)+\exp(\beta\Delta A_3-\beta\Delta A_2)}
\end{aligned} \tag{3.3}$$

This means that knowing the variables ΔA_2 and ΔA_3 is enough for the traders to use Eq. (3.2) to choose the market they want to trade with. Then, we perform a Kramers-Moyal expansion of the traders learning dynamics and obtain two Fokker-Planck equations (one for each class $c \in \{1,2\}$ of traders) for the distribution of preferences differences: $\mathbb{P}(\mathbf{\Delta A}^{(c)}, t)$:

$$\begin{aligned}
\partial_t \mathbb{P}(\mathbf{\Delta A}^{(c)}, t) = &- \sum_{m=2}^{3} \partial_{\Delta A_m^{(c)}} \left[\mu_m^{(c)}(\mathbf{\Delta A}^{(c)}, f_1, f_2, f_3) \mathbb{P}(\mathbf{\Delta A}^{(c)}, t) \right] \\
&+ \frac{r}{2} \sum_{m,m'=2}^{3} \partial_{\Delta A_m^{(c)}} \partial_{\Delta A_{m'}^{(c)}} \left[\Sigma_{mm'}^{(c)}(\mathbf{\Delta A}^{(c)}, f_1, f_2, f_3) \mathbb{P}(\mathbf{\Delta A}^{(c)}, t) \right]
\end{aligned} \tag{3.4}$$

where $\mathbf{\Delta A}^{(c)} = (\Delta A_2^{(c)}, \Delta A_3^{(c)})$ and f_m is the ratio of buyers over sellers at market m. The expressions for the drift vector $\mu_m^{(c)}(\mathbf{\Delta A}^{(c)}, f_1, f_2, f_3)$ and the covariance matrix $\Sigma_{mm'}^{(c)}(\mathbf{\Delta A}^{(c)}, f_1, f_2, f_3)$ are given in Appendix 3.A. We can derive a Fokker Planck equation for the two score differences only and not for the three variables $A_1^{(c)}, A_2^{(c)}, A_3^{(c)}$ because we restrict our analysis to traders learning using fictitious play ($\alpha = 1$). For $\alpha < 1$ we would have a 3-dimensional Fokker Planck equation instead of the simpler 2-D Fokker equation above, which would complicate our analysis.

3.2.2 Comparison with multi-agent simulations

To get an idea of the different types of attraction distributions in this extension of the double auction markets model, we run multi-agent simulations for various intensities of choice and market biases. We show an example of

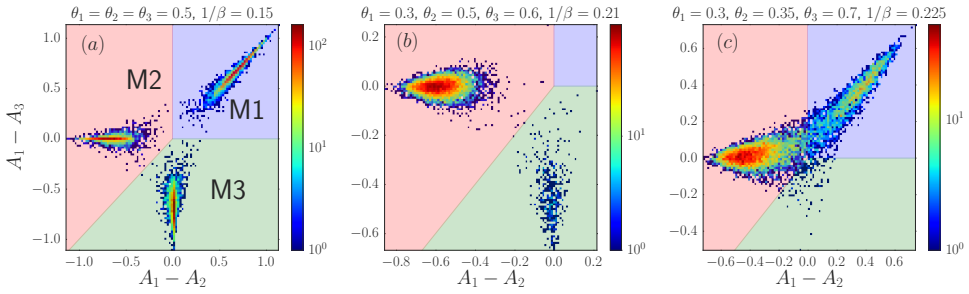

Figure 3.1: Log-distribution of scores of traders for parameters set as indicated in each graph title. In (a), the population is segregated into three groups. In (b), the distribution has two peaks: one large peak and one peak that becomes exponentially small as the memory length decreases. In (c), the distribution of scores has two large peaks. To obtain those graphs, we ran simulations with $r = 0.01$ and $N/2 = 10,000$ traders in each class, until a steady state was reached. Traders from class 1 have preference to buy $p_{\mathrm{b}}^{(1)} = 0.2$ and traders from class 2 have preference to buy $p_{\mathrm{b}}^{(2)} = 0.8$. To ease the interpretation of the attraction distributions, the $(A_1 - A_2, A_1 - A_3)$ plane is shown subdivided into three zones according to which market agents with the corresponding preferences choose most of the time. The zones are coloured blue, red and green for the markets 1, 2 and 3, respectively, as indicated in (a).

three qualitatively different distributions obtained for different market biases in Fig. 3.1. Also represented is the market preferred by a trader depending on his preferences in each of the panels of Fig. 3.1, to facilitate the interpretation of the scores distributions. We now give a brief description of the scores distributions in each of the panels and explain the difference between (i) strong segregation, which persists in the large memory limit, and (ii) weak segregation, which disappear in the large memory limit. In panel (a) of Fig. 3.1, one sees that the distribution of scores has three peaks, all of which have a size of order $O(1)$ and correspond to traders who choose to trade mainly at a single market. In other words, the population splits into three subpopulations of traders that are "attached" to one market over the others. Such distributions of scores with more than one peak with a size of order one are called *strongly segregated*. The

second distribution, shown in panel (b), corresponds to a population divided into two subpopulations with different sizes: the first subpopulation with a size of order N goes to the second market and the second one whose size decreases exponentially as $r \to 0$ prefers to go to the first market. Such attraction distribution with strictly more than one peak but with only one peak having a weight of order 1 are said to be *weakly segregated*. The distribution plotted in panel (c) corresponds to a strongly segregated population part of which goes to the second market and the rest to the first market.

From the example above we see that there is a variety of qualitatively different structures of the attraction distributions (number of peaks, size). This gives a first taste of the rich behaviour of steady states of the learning dynamics for traders choosing between more than two markets (see Fig. 3.1). As in chapter 2, we will study the emergence of segregation in the large memory limit ($r \to 0$) for various ranges of parameters and use our results to construct a simple intuitive hypothesis for the possible causes of segregation.

3.3 Three fair markets

We start by looking at what happens when the three markets available are all fair, *i.e.* $\theta_1 = \theta_2 = \theta_3 = 0.5$. This means they set their trading price to be exactly the mean of the average bid and the average ask. Here, we focus on the steady state of traders' learning dynamics without symmetry breaking (this is the one we observe in multiagent simulations). Since the three markets have the same bias θ, they have the same number of buyers (resp. sellers) and because the classes have symmetric preferences to buy $p_{\mathrm{b}}^{(1)} = 1 - p_{\mathrm{b}}^{(2)}$, the difference between the number of buyers and the number of sellers is of order $\sqrt{N}$, $N_{\mathrm{b}} = N_{\mathrm{s}} + O(\sqrt{N})$. As a consequence, in the large size limit, the ratio of the number of buyers to the number of sellers in each market is equal to

1. This simplification is the reason why we choose to start this chapter with this simple case, which is a good way to get familiar with the phenomenon of segregation across three double auction markets.

Since we do not need to determine the aggregates by self-consistency as was the case in Chap. 2 we start by looking at the fixed point structure of a single agent dynamics when the intensity of choice is small. As expected, the only fixed point of the learning dynamics is $A_1 - A_2 = A_1 - A_3 = 0$ and corresponds to a trader who chooses to randomise totally between the three markets (see Fig. 3.2(a)). When the intensity of choice β reaches a critical value $\beta_c = 1/0.254$ three saddle node bifurcations take place simultaneously and three pairs of stable and unstable fixed points appear (see Fig. 3.2(b)). The reason why those three saddle node bifurcations take place at the same time is that all the markets have the same bias $\theta = 0.5$. In the more general case where the three markets are different, we expect the appearance of the fixed points to take place at different values of β. As we continue to increase the intensity of choice there is a small range, $1/0.252 \geq \beta \geq 1/0.254$, where the system is weakly segregated because the central fixed point is stable and corresponds to a finite peak in the distribution of preferences of traders from population 1 while the three outer meta-stable fixed points correspond to small peaks in the attraction distribution (*i.e.* peaks that decay exponentially with the memory length $1/r$). Then at the second critical value of β: $\beta'_c = 1/0.252$ the three outer fixed points become stable and the system undergoes a strong segregation transition. For any values of β above this second segregation threshold, the system will be strongly segregated as the distribution of preferences of the traders will have three peaks of equal weight, each of which corresponds to a stable fixed points of the single agents dynamics (red points in Fig. 3.2(c,d)). For $1/0.237 \leq \beta \leq 1/0.252$, the distribution of preferences will still have an additional peak at the meta-stable fixed point located in $(0, 0)$ but the weight of this peak will become exponentially small as the memory length increases (see Fig. 3.2(c)). This meta-stable fixed point and the associated small peak in the

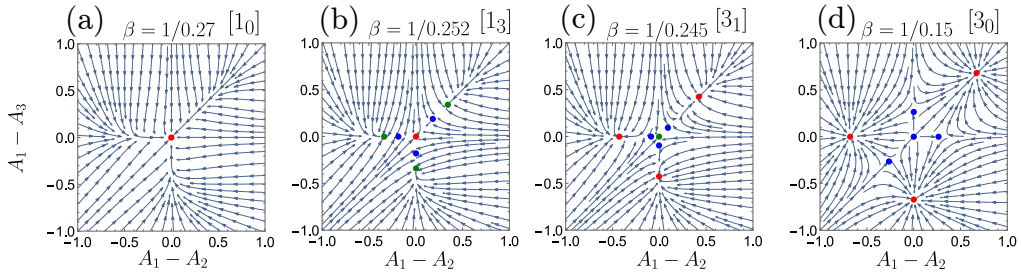

Figure 3.2: Flow diagram and fixed points of the learning dynamics of a single trader with $p_{\mathrm{b}}^{(1)} = 0.2$. Below the weak segregation threshold $\beta = 1/0.254$, the dynamics has only one fixed point which is stable (red) (see Panel (a)). When β reaches the weak segregation threshold $\beta_{\mathrm{c}} = 1/0.254$, three pairs of unstable (blue) and meta-stable (green) fixed points appear and the system becomes weakly segregated with one large peak which corresponds to traders randomizing between the three markets and three small peaks each of which corresponds to traders trading mainly at only one of the three available markets (panel (b)). At $\beta'_{\mathrm{c}} = 1/0.252$, the three outer fixed points become stable and the central one meta-stable and the system is now strongly segregated with three peaks of equal size each of which corresponds to the strategy to trade at only one of the market (panel (c)). Then as β increases, the meta-stable fixed points becomes unstable as shown in panel (d). Above each graph, after the value of β, we indicate in brackets the category to which each of the fixed points structure belongs (see main text for details).

attraction distribution then disappear for $\beta \geq \beta''_{\mathrm{c}} = 1/0.237$ (see Fig. 3.2(d))

When looking at the deterministic dynamics for low intensity of choice (see Fig. 3.2(a)) it is obvious that the system is not segregated (the deterministic dynamics have only one stable fixed point) and for large intensity of choice, we see that the system is strongly segregated with three peaks of equal size (see Fig. 3.2(d)). When $1/0.237 > \beta > 1/0.254$ the fixed point structure

values of β	Category for pop.1	Category for pop.2
$\beta \in (0, 1/0.254]$	1_0	1_0
$\beta \in [1/0.254, 1/0.252]$	1_3	1_3
$\beta \in [1/0.252, 1/0.237]$	3_1	3_1
$\beta > 1/0.237$	3_0	3_0

Table 3.1: Zone of existence of each of the categories of scores distribution, when traders with symmetric fixed buy/sell preferences $p_{\rm b}^{(1)} = 1 - p_{\rm b}^{(2)} = 0.2$ choose between three fair markets

looks as shown in Fig. 3.2(b,c) and knowing the deterministic dynamics is not sufficient to distinguish between stable and meta-stable fixed points. To assess the stability or otherwise of fixed points in Fig. 3.2, we use the Freidlin-Wentzell approach detailed in Sec. 2.5. For $1/0.237 > \beta > 1/0.254$, as the three markets have the same bias θ, either all the outer fixed points are stable, or they are all meta-stable. Hence we can restrict our analysis to only one of the outer fixed points and the central fixed point. Let us take the top right fixed point and the central fixed point $A_1 - A_2 = A_1 - A_3 = 0$. We calculate the transition rates between those two fixed points with the Freidlin-Wentzell theory and assess the stability or otherwise of the outer fixed point. Since the rates are exponential in the relevant action, the stable state is the one with the larger action.

As explained in the discussion above, traders' learning dynamics has a rich variety of qualitatively different steady states. To categorise these steady states, we will focus on the number of peaks in the attraction distribution and on whether they persist or not in the large memory limit. As we mention in Chap. 2 for large memory (small r), the distribution of preference differences is composed of peaks of width proportional to the inverse memory length r. Those can be classified into two categories: (i) the peaks whose size is of order one (they are located on stable fixed points of the single agent dynamics) which

we call *large peaks*; (ii) the peaks whose weight decay exponentially with the memory length $1/r$ which we call *small peaks*. In this paragraph we introduce a classification of such distribution depending on the number of *large* and *small* peaks. The number of small ($n_{\mathrm{s}}^{(c)}$) and large ($n_{\mathrm{L}}^{(c)}$) peaks of a class c are the quantities we will use to categorise a attraction distribution and we will label the corresponding category by $(n_{\mathrm{L}}^{(1)}{}_{n_{\mathrm{s}}^{(1)}}, n_{\mathrm{L}}^{(2)}{}_{n_{\mathrm{s}}^{(2)}})$. For example the distribution which corresponds to the fixed point structure in Fig. 3.2(b) is labeled $(1_3\ 1_3)$ as the distribution of attraction differences of both classes of traders has one large peaks and three small peaks. To complete this explanation, I wrote the category of each of the fixed points structures in Fig. 3.2 in the top right of each plot [in brackets].

In this first analysis of segregation when traders can choose between three fair double auction markets, we observed a rich variety of qualitatively different preferences distributions in the steady state of the traders' learning dynamics, some of which are segregated. One question of interest is whether this segregation comes from the fact that all the markets are identical or if this phenomenon will persist for markets with different biases. To answer this question, in the next section we extend our analysis to markets with different biases.

3.4 Exploration of the parameter space: markets with different biases

In this section, we are interested in exploring the different combinations of markets biases θ_1, θ_2, θ_3 and their influence on the phenomenon of segregation. Each market bias takes values between zero and one, *i.e.* the parameters space we explore is a cube with edges of length one. Of course the phenomenon of segregation is independent under permutation of the market biases as this

effectively just changes the labelling of the markets. We can therefore restrict our analysis to 1/6 of the cube where $\theta_1 \leq \theta_2 \leq \theta_3$ and can reconstruct the behaviour in the rest of the parameter space by symmetry. Although this reasoning provides an intuitive explanation on how to restrict the parameter space to study, we will not strictly respect this ordering in our parameter exploration, in order to get simpler phase diagrams. Since phase diagrams are easier to plot in 2-D (one dimension for the bias and one dimension for the inverse intensity of choice) we are going to fix two of the three market biases and vary only one of them in three different manners: (i) $\theta_2 = 0.5$ and $\theta_1 = 1 - \theta_3$ and we vary θ_1 as a free parameter between zero and 1/2, (ii) $\theta_1 = 0.3$, $\theta_2 = 0.7$ and θ_3 is a free parameter, (iii) $\theta_1 = 0.3$, $\theta_2 = 0.5$ and θ_3 is a free parameter that we vary between zero and one. As will become clear in the rest of this section, the parameters studied here allow for the analysis of properties such as the market symmetry, the "distance" between market biases as well as the effect of market fairness on the occurrence of segregation.

3.4.1 Two symmetric markets and one fair market

In this subsection, we study segregation in a system with two symmetric markets and one fair market. As in our study of three fair markets, for simplicity, we restrict our study to solutions that do not break the market symmetries as they are the ones that we see in our multi-agent simulations. As in the case of three fair markets detailed in Sec. 3.3 we will use the symmetries of the problem to restrict the possible values of the aggregates. In particular, one can show that the ratio of buyers over sellers in market one is the inverse of the same ratio in market three, and that the ratio of buyers over sellers at the fair market is unity.

One first note that when $\theta_1 = 1 - \theta_3$ and $\theta_2 = 0.5$, for traders with symmetric preferences to buy, the role played by market one for traders from class 1 is the

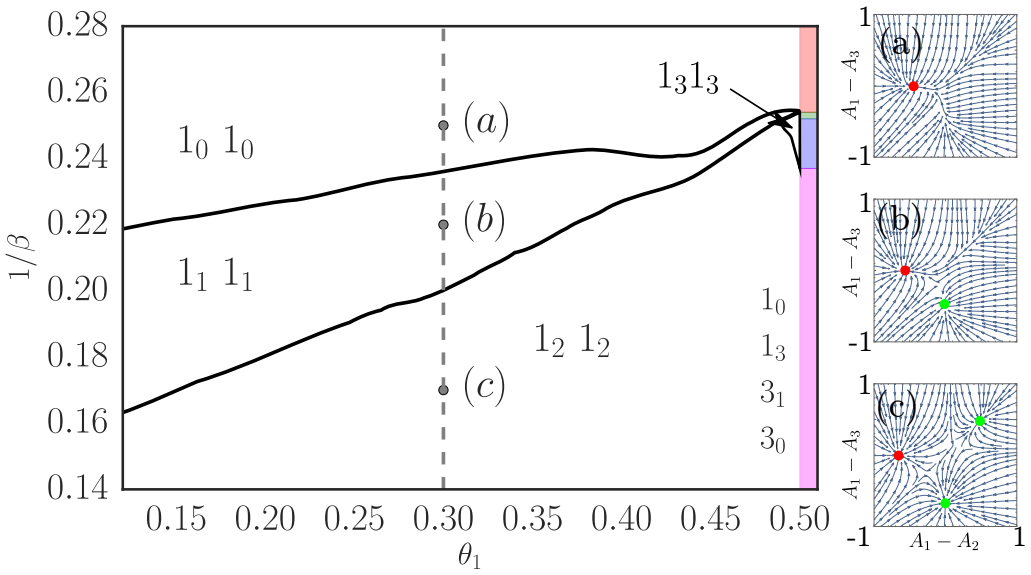

Figure 3.3: Peak structure of the steady state distribution of traders' preferences when they learn to choose between three markets, two of which have symmetric market biases $\theta_1 = 1-\theta_3$ and one of which is fair. The three insets on the right show the fixed point structure for an agent from class 1, for $\theta_1 = 0.3$ and different β as indicated in the phase diagram by grey points. In the inset, red points correspond to a stable (large peak) and green points to a metastable fixed point (small peak). The color band at $\theta_1 = 0.5$ shows the type of attraction distribution when $\theta_1 = 0.5$, *i.e.* when the three markets are fair (see Tab. 3.1). Note that the zones $(1_3\ 1_3)$ and $(1_1\ 1_1)$ are separated by a a a small "finger" of the $(1_2; 1_2)$ zone that we do not display here to keep the phase diagram clear.

same as the role played by market three for traders from class 2 and vice-versa. As a consequence, the probability of trading at the first market for a trader from class 1 (resp. 2) is equal to the probability of trading at the third market for a trader of class 2 (resp. 1). Hence using the symmetry of the model, we

can write the buyer/seller ratios in market 1 and 3 as

$$f_1 = \frac{\mathbb{P}^{(1)}(M=1)p_{\mathrm{b}}^{(1)} + \mathbb{P}^{(2)}(M=1)p_{\mathrm{b}}^{(2)}}{\mathbb{P}^{(1)}(M=1)(1-p_{\mathrm{b}}^{(1)}) + \mathbb{P}^{(2)}(M=1)(1-p_{\mathrm{b}}^{(2)})} \tag{3.5}$$

$$f_2 = \frac{\mathbb{P}^{(1)}(M=3)p_{\mathrm{b}}^{(1)} + \mathbb{P}^{(2)}(M=3)p_{\mathrm{b}}^{(2)}}{\mathbb{P}^{(1)}(M=3)(1-p_{\mathrm{b}}^{(1)}) + \mathbb{P}^{(2)}(M=3)(1-p_{\mathrm{b}}^{(2)})} \tag{3.6}$$

When substituting into these expressions the equalities $\mathbb{P}^{(1)}(M=1) = \mathbb{P}^{(2)}(M=3)$, $\mathbb{P}^{(2)}(M=1) = \mathbb{P}^{(1)}(M=3)$ and remembering that $p_{\mathrm{b}}^{(1)} = 1 - p_{\mathrm{b}}^{(2)}$, one sees that $f_1 = 1/f_3$. With the same reasoning, one can show that the ratio of buyers over sellers at the fair market (market number two) is 1.

Let us first calculate the value of the intensity of choice at which traders start to segregate weakly. To do so, for a given value of the free parameter θ_1, we start from low values of β and gradually increase the intensity of choice until it reaches a critical value where the single agent dynamics has two stable fixed points. Those values of β are shown by the upper solid line in Fig. 3.3. The natural continuation of this analysis is to look – if it exists – for the strong segregation threshold. While thanks to our previous analysis of symmetric markets we know that for $\theta_1 = 0.5$ strong segregation takes place at $\beta = 1/0.252$, our numerical methods show that for markets reasonably asymmetric *i.e.* $\theta_1 < 0.48$, strong segregation does not take place in the range of values of β we look at in our phase diagram. For θ_1 between 0.48 and 0.5, our numerics suggest possible strong segregation but within the precision of our action minimization we could not definitely conclude whether it occurs or not. The absence of segregation for $\theta \neq 0.5$ is the reason why the 1_3 and 3_1 zones that exist when the three markets are fair ($\theta_1 = 0.5$) merge into a 1_3 zone for values of θ_1 different from 0.5. Note that the range of values of θ_1 for which this 1_3 solution exists is rather small (see Fig. 3.3), because the central stable fixed point in the 1_3 state rapidly becomes unstable as θ decreases. In Fig. 3.3 we see that for any value of β, the majority of the traders will prefer to trade at the fair market (market number two). However, as the intensity of choice

increases, the two new peaks which appear correspond to going to trade at market number one or at the market number three.

In this subsection, we discussed weak segregation across three double auction markets when two of them are symmetric, and the third one is fair. Surprisingly, apart from the particular case when the three markets are all fair, *strong* segregation does not take place. This raises the question of whether the absence of strong segregation is a general feature of traders choosing between three market, *i.e.* if it can arise only due to the symmetries between the markets. This motivates the exploration of further parameter settings in Secs. 3.4.2 and 3.4.3

3.4.2 Two symmetric markets and one biased market

In the previous section, we discussed the fact that for two symmetric markets and one fair market only weak segregation takes place. In this part, we consider two symmetric markets with fixed market biases: $\theta_1 = 0.3$ and $\theta_3 = 0.7$. For these fixed parameters we show in Fig. 3.4 a phase diagram of the type of attraction distribution as a function of the bias of the second market $\theta_2 \in [0.3, 0.7]$.

We first note that strong segregation appears, and does so across a reasonably broad range of market biases. Nonetheless, the range of market biases for which strong segregation takes place (grey zone in Fig. 3.4) is limited. Indeed, segregation occurs only for $\theta_2 \notin [0.45, 0.55]$ *i.e.* when the second market is sufficiently biased. For $\theta_2 < 0.45$ (resp. $\theta_2 > 0.55$) the traders from the second (resp. first) class *strongly segregate* across the two markets that maximise average profit per trade for each class. One interesting research question for future work is the shape of the phase diagram below the first strong segregation threshold (grey zone in Fig. 3.4). We choose not to explore such values of β

here as this would require the numerical solution of self-consistency conditions for multiple aggregates in the presence of two (or more) strong segregation peaks in the traders' attraction distributions, and leave this question for future work. However, it is possible to get an intuition about the shape of the phase diagram below this threshold by extrapolating the zones of weak segregation when the second market is close to fair (dotted lines in Fig. 3.4). Another possibility to get an idea of the shape of the phase diagram is to run multi-agent simulations and plot the distribution of scores (this is what we do in Fig. 3.5).

Having described the range of values of θ_2 for which strong segregation takes place, we inspect more closely the range of parameters for which only weak segregation occurs. To do so, we look at how the attraction distributions of both classes of traders evolve at fixed $\theta_2 = 0.47$ when β increases. For values of β small enough in relation to the agents' preferences, they will essentially randomise. As β increases, the traders from the two classes synchronise at the market that is close to fair (market two) and provides a good trade-off between profit and trading volume. Then, as β increases, small peaks arise in the distribution of scores of each of the classes but most of the traders remain in the fairer market. At $\beta = 1/0.246$ a peak corresponding to the strategy "trading at the profit maximizing market" (market one, which has $\theta_1 = 0.3$) appears in the attraction distribution of the agents from the second class. Then at $\beta = 1/0.228$, a peak corresponding to the strategy "trading at the profit maximizing market" (market three with $\theta_3 = 0.7$) appears in the attraction distribution of the agents from the first class. After those two successive appearances of weak segregation between the fair market and the profit maximizing market for both class 1 and class 2, a new peak in the attraction distribution – which corresponds to the strategy "trading at the volume maximizing market" – appears successively for class 1 at $\beta = 1/0.207$ and then for class 2 at $\beta = 1/0.198$.

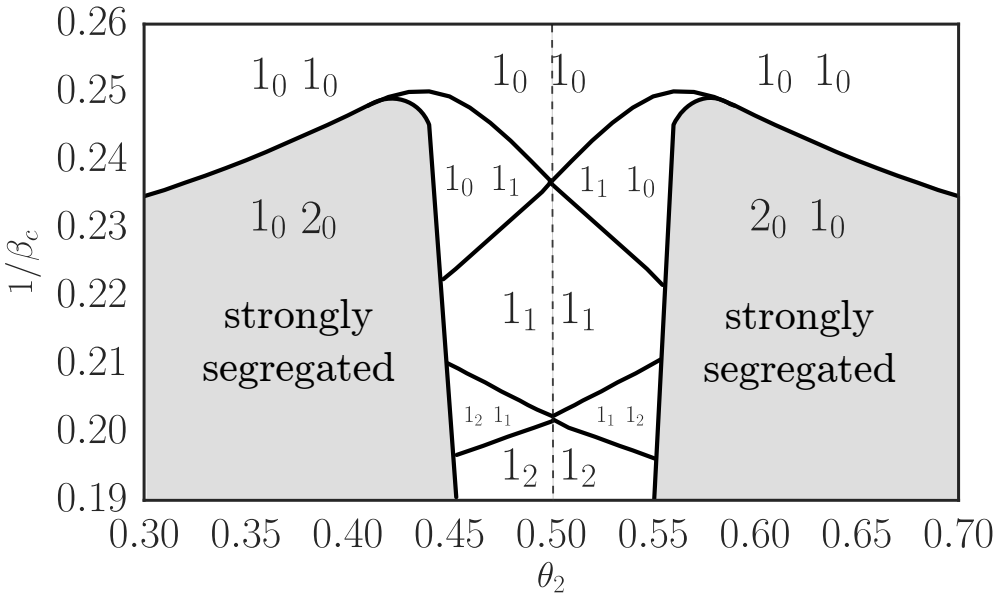

Figure 3.4: Types of attraction distributions in the steady state of EWA learning dynamics for markets with biases $\theta_1 = 1 - \theta_3 = 0.3$ and varying θ_2. The grey zone indicates the region in parameter space where the distribution of scores has two peaks of large size, *i.e.* where strong segregation occurs. Note that between an unsegregated region ($1_0\ 1_0$) and a strongly segregated (($2_0\ 1_0$) resp. ($1_0\ 2_0$)) there is always a weakly segregated region (($1_1\ 1_0$) resp. ($1_0\ 1_1$)) but these are mostly too narrow to be visible in this phase diagram. There are likely to be further phase boundaries within the strongly segregated region as indicated by the dotted lines, but we have not calculated those in detail. The dashed line in the centre corresponds to the dashed line in Fig. 3.3.

Our phase diagram suggests that fairness of the second market weakens segregation. However, in the phase diagram of Fig. 3.4 we checked the onset of segregation only for reasonably large inverse decision strengths (*i.e.* $1/\beta > 0.19$). It is therefore still possible that strong segregation might occur at sufficiently high intensity of choice $1/\beta < 0/19$ even in the range $\theta \in [0.45, 0.55]$. As the intensity of choice increases, each class will first weakly segregate across

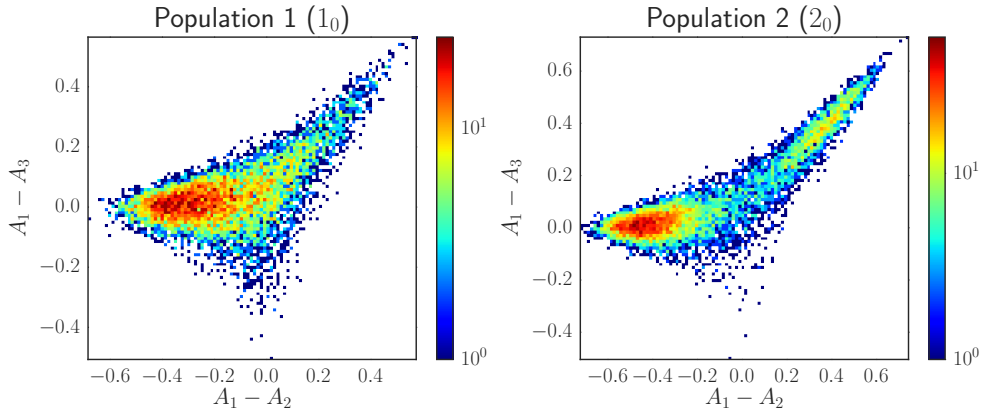

Figure 3.5: Log-distribution of score differences of traders who choose between three markets with market biases $(\theta_1, \theta_2, \theta_3) = (0.3, 0.35, 0.7)$. The population consists of two classes of $N/2 = 10^4$ traders with symmetric buy-sell preferences $p_{\mathrm{b}}^{(1)} = 1 - p_{\mathrm{b}}^{(2)} = 0.2$, inverse memory length $r = 0.01$ and intensity of choice $\beta = 1/0.21$. We see that the attraction distribution of the second population is strongly segregated so that the state overall is of type $(1_0\ 2_0)$ in our notation, as predicted in the phase diagram of Fig. 3.4.

the second market (close to fair) and the profit maximizing market and then weakly segregate across the three markets. On the other hand, if the second market is not fair the population for which this market is profit-oriented will strongly segregate between the two profit maximizing markets and the other class will only trade at the fair market. The results of this subsection suggest that as soon as traders have at their disposal a reasonably fair market, they are not going to segregate and will prefer to trade with the fair market; when they have no fair market they will always prefer the profit maximizing market. To put it differently, we expect that the preferences of a trader among the different types of markets are generally ordered as follows:

market close to fair > profit oriented market > volume oriented market

In the next section, we relax the assumption that two of the three available markets are symmetric, force one of the markets to be fair and investigate the emergence or otherwise of strong and weak segregation.

3.4.3 Markets without symmetry

The two examples presented in subsections 3.4.2 and 3.4.1 lead to the conjecture that the presence of a fair market – which provides a good trade-off between trades profit and trades volume – can suppress segregation. To confirm this conjecture, we consider three markets where the first one is biased toward buyers ($\theta_1 = 0.3$) and the second one is fair ($\theta_2 = 0.5$); the bias of the third market is the parameter we will vary. As we did in the previous subsections, we will draw a phase diagram of the type of distribution of scores of the traders depending on the intensity of choice β and the bias of the third market $\theta_3 \in [0, 1]$.

In the phase diagram in Fig. 3.6 one can see that for the range of parameters we explore, the distributions of scores of both of the classes of traders are only have one peak whose size is of order one, *i.e.* depending on the value of β and θ_3 they are unsegregated or weakly segregated with – in the latter case – one or two small peaks. In Fig. 3.6, one sees that at fixed θ_3, once the intensity of choice goes above a certain threshold value shown by the full green line in Fig. 3.6, a weak peak corresponding to the strategy "trading at market one" appears in the distribution of scores of the second class of agents. Then, when β is above the second segregation threshold (represented by the green dotted line in Fig. 3.6), the same type of weak peak appears in the distribution of scores of the first class of agents. The fact that the two green lines just described are close to horizontal reflects the fact that since almost all of the population trades at the fair market, the bias of the third market will not influence the preference of traders. This is the reason why the critical intensity of choice at which traders of class 1 (resp. class 2) will weakly segregate between market 1 and market 2 is almost *independent on the third market's bias*. The same is not true of the threshold for the appearance of a peak corresponding to the strategy "trade at market 3": the red lines in Fig. 3.6 showing this threshold

are inclined rather than horizontal. This is the sign that the appearance of a weak peak corresponding to the strategy "trading at market three" depends on both the intensity of choice β and on the bias of the third market θ_3. Our numerics did not show strong segregation for inverse intensity of choice higher than 0.17, moreover, extrapolating our current results, we do not expect strong segregation to take place even for $1/\beta < 0.17$.

3.5 Discussion/Conclusion

In this chapter we used the equal action formalism developed in chapter 2 to study strong segregation of traders with fixed buy/sell preferences across three markets. After setting out the motivation for the extension to segregation across three markets in Sec. 3.1, we described precisely how we extend the double auction markets model from two to three markets. Motivated by the wide variety of structures of the attraction distributions that one observes in multi-agent simulations, we explored in Secs. 3.3 and 3.4 different combinations of market biases and their influence on the phenomenon of segregation. First, in Sec. 3.3, we studied segregation across three fair markets, *i.e.* with $\theta_1 = \theta_2 = \theta_3 = 0.5$; this is the only set of market biases for which we found segregation of both of classes of traders into three groups. In Sec. 3.4 we explored different market configuration to get an intuition of the factors that drive segregation. This enabled us to identify two principal causes of segregation (i) *the closeness between the markets*, (ii) *the average volume of trade and profit earned in this market*. The *closeness* between two markets is going to enhance segregation because traders are more likely to split across two markets if they cannot tell them apart. This effect is visible in Sec. 3.4.3 where the strong and weak segregation thresholds are the highest when the second market and the fair market have the same bias. The ordering of the appearance of the peaks in the traders' scores distribution suggests – as we pointed out in Sec. 3.4.2 – that

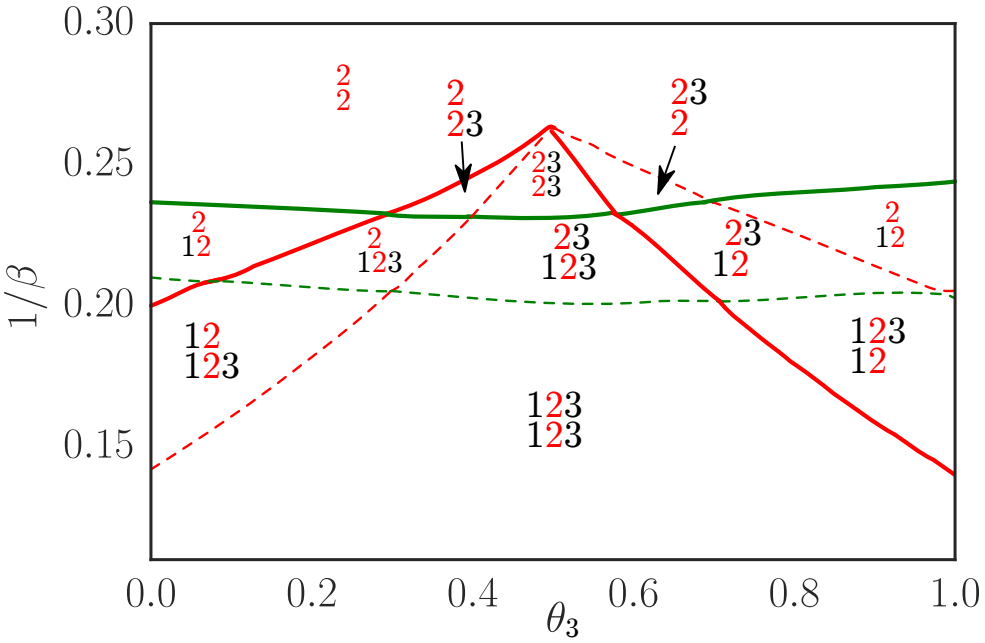

Figure 3.6: Peak structure of the different attraction distributions when $\theta_1 = 0.3$, $\theta_2 = 0.5$, $p_{\mathrm{b}}^{(1)} = 1 - p_{\mathrm{b}}^{(2)} = 0.2$. We label each zone by two sequences of numbers: the top one corresponds to class one and the bottom one to class 2. The red number correspond to the favourite market in the large peak of the attraction distribution and the black one to the favourite market in the weak peaks of the attraction distribution. For example the bottom zone $\begin{smallmatrix}123\\23\end{smallmatrix}$ corresponds to a $(1_2\ 1_1)$ distribution of scores where both of the classes have one main peak at the fair market (market two), class one has two weak peaks at market two and market three and class two has a single weak peak at market three.

traders will have preference for markets that provide a good balance between volume of trades and profit of trades, then as the intensity of choice decreases they will go first to the one which maximises their profit and then subsequently to the one which maximises their volume of trades.

We examined the existence of strong segregation only for moderate intensities of choice *i.e.* $1/\beta > 0.2$. The reason for this restriction is that the minimal action algorithms sometimes fails to find a physical minimal action path for large β. For this reason working on the robustness and accuracy of the action minimization algorithm would be an obvious continuation of this work. We could for example use the *geometric minimum action method* [48] in which the limit $t_1 \to \infty$ and $t_2 \to \infty$ is assumed. This makes it possible to write the action of a path as an integral over a normalised arc-length $\phi \in [0, 1]$ rather than an integral over a time $t \in [-\infty, \infty]$.

A further natural extension of this work would be to increase the number of markets. Such an extension is likely to be challenging for two reasons: (i) the parameter space of market biases would be of dimension n (ii) minimal action paths would have to be found in a space of attraction differences of dimension $n - 1$. On the plus side, tackling these issues would enable one to check how general the results we obtained for segregation across two markets are.

Finally, it would also be interesting to find a robust way to assess whether a distribution of scores is segregated or otherwise *from simulation data* to locate more precisely the strong segregation region in Fig. 3.6. The Binder cumulant [6] fails to predict segregation when the peaks of the attraction distribution do not have the same size. Alternatively, to assess the bimodality of the scores distributions one can use *Hartigan's DIP test* [47], which uses the distance between the distribution of scores and the "best fitting" unimodal distribution to assess the unimodality of the attraction distribution or otherwise. The drawback of this test compared to the Binder cumulant is that it only

gives information about the shape of the distribution when traders have finite memory but its predictions do not generalise to the limit $r \to 0$, which is the limit we are interested in.

Appendix

3.A Kramer-Moyals expansion

In this appendix we give the expression of the drift and covariance matrix which appear in the Kramer-Moyals expansion in Eq. (3.4). The step by step method we use for the full derivations of the drift and covariance matrix are detailed in the thesis of Alorić [1]. First, the expression of the drifts are:

$$\mu_2^{(c)}(\mathbf{\Delta A}^{(c)}, f_1, f_2, f_3) = \Big(\mathcal{P}_1^{(c)}(f_1)\mathbb{P}(M=1) - \mathcal{P}_2^{(c)}(f_2)\mathbb{P}(M=2)\Big) - \Delta A_2^{(c)} \tag{3.7}$$

$$\mu_3^{(c)}(\mathbf{\Delta A}^{(c)}, f_1, f_2, f_3) = \Big(\mathcal{P}_1^{(c)}(f_1)\mathbb{P}(M=1) - \mathcal{P}_3^{(c)}(f_3)\mathbb{P}(M=3)\Big) - \Delta A_3^{(c)} \tag{3.8}$$

$\mathcal{P}_m^{(c)}(f_m)$ is the average payoff of a trader from class c at marker m and $\mathbb{P}(M = m)$ is the probability to trade at market m which depends on the scores differences vector $\mathbf{\Delta} A^{(c)}$ defined in Eq. (3.3). We drop the dependence on the score differences deliberately to lighten the notations. In order to check the validity of our calculations we compared the dynamics of the aggregate f_1 during a multiagent simulation with the evolution of the aggregates under the homogeneous population dynamics as which is detailed in Sec. 2.4.2 both of which are consistent with each other as shown in Fig. 3.A.1. We now look at

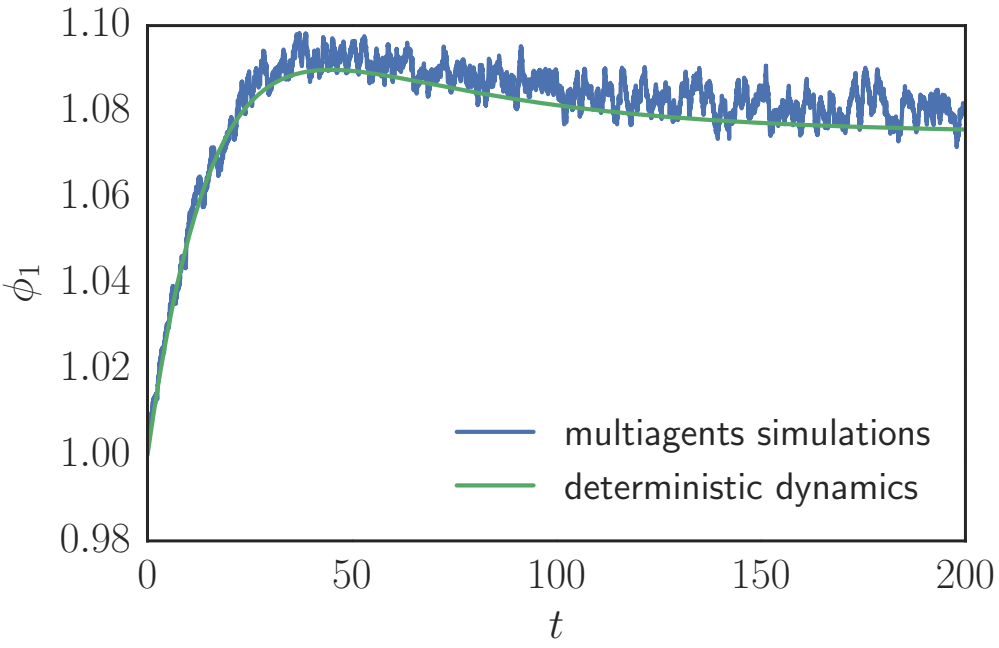

Figure 3.A.1: Comparison between the time series of the aggregates at the first market during a multiagents simulations ($r = 0.01$ and 10^4 agents in each class) and their evolution under the homogeneous population dynamics. The parameters for the plots in this figure are $(\theta_1, \theta_2, \theta_3) = (0.2, 0.5, 0.8)$, $\beta = 1/0.3$ and $p_{\mathrm{b}}^{(1)} = 1 - p_{\mathrm{b}}^{(2)} = 0.2$

the covarianc matrix

$$\begin{pmatrix} \Sigma_{22}^{(c)} & \Sigma_{23}^{(c)} \\ \Sigma_{23}^{(c)} & \Sigma_{33}^{(c)} \end{pmatrix} \tag{3.9}$$

whose expression is indicated below:

$$\begin{aligned}\Sigma_{22}^{(c)}(\mathbf{\Delta A}^{(c)}, f_1, f_2, f_3) = &\left(\mathcal{Q}_1^{(c)}(f_1) - 2\Delta A_2^{(c)}\mathcal{P}_1^{(c)}(f_1)\right)\mathbb{P}(M=1)\\ &+\left(\mathcal{Q}_2^{(c)}(f_2) - 2\Delta A_2^{(c)}\mathcal{P}_2^{(c)}(f_2)\right)\mathbb{P}(M=2)\\ &+\Delta {A_2^{(c)}}^2 \qquad (3.10)\\ \Sigma_{33}^{(c)}(\mathbf{\Delta A}^{(c)}, f_1, f_2, f_3) = &\left(\mathcal{Q}_1^{(c)}(f_1) - 2\Delta A_3^{(c)}\mathcal{P}_1^{(c)}(f_1)\right)\mathbb{P}(M=1)\\ &+\left(\mathcal{Q}_3^{(c)}(f_3) - 2\Delta A_3^{(c)}\mathcal{P}_3^{(c)}(f_3)\right)\mathbb{P}(M=3)\\ &+\Delta {A_3^{(c)}}^2 \qquad (3.11)\\ \Sigma_{23}^{(c)}(\mathbf{\Delta A}^{(c)}, f_1, f_2, f_3) = &\Delta A_2^{(c)}\left(\mathbb{P}(M=3)\mathcal{P}_3^{(c)}(f_3) - \mathbb{P}(M=1)\mathcal{P}_1^{(c)}(f_1)\right)\\ &+\Delta A_3^{(c)}\left(\mathbb{P}(M=2)\mathcal{P}_2^{(c)}(f_2) - \mathbb{P}(M=1)\mathcal{P}_1^{(c)}(f_1)\right)\\ &+\mathbb{P}(M=1)\mathcal{Q}_1^{(c)}(f_1) + \Delta A_2^{(c)}\Delta A_3^{(c)} \qquad (3.12)\end{aligned}$$

where $\mathcal{Q}_m^{(c)}(f_m)$ is defined in Eq. (2.36)

CHAPTER 4

Varying memory and decision strength

4.1 Introduction

In Chap. 2 and 3 we made the assumption that the traders discounted the payoffs of past trades at the same rate r. Although this assumption made our analysis clearer, it ignores the diversity of traders' objectives. In practice, traders have access to different pieces of information, have different expectations, and can have different interpretations of shared information that is widely available. For example, the time scale over which *investors* expect to make a profit is much larger than the timescale over which *speculators* expect to make a profit. This chapter aims to present a simple extension of the model defined in Chap. 2 which takes into account heterogeneity in the behaviour of traders.

The consequences of heterogeneity in the information received by traders is a widely discussed topic. For example, in Ref. [98], subtitled "The luck of the uninformed", Toth *et al.* study the dynamics of traders with different information levels who learn to trade in a double auction market. Surprisingly in such settings, the profit of traders doesn't always increase with their information

level. Bloembergen *et al.* explain in [8]: "One possible theory explaining this phenomenon is that more information helps during trends, whereas limited knowledge may be erroneous when the trend reverses; uninformed traders are safe from these systematic mistakes". The study of the dynamics of traders with heterogeneous memory playing the minority game has also attracted attention from physicists [106, 52]. For example, [52] reports enhanced winnings for traders with long memory in a population where the memory length of players is heterogeneous. Our aim is to investigate the robustness of segregation to the introduction of heterogeneities in both the memory and the intensity of choice of traders.

This chapter is organised as follows: in Sec. 4.2 we introduce the theoretical formalism we use in order to study the co-learning between fast and slow traders. In Sec. 4.3 we apply the theory developed in Sec. 4.2 to investigate the influence of heterogeneity of traders on the phenomenon of segregation.

4.2 Dynamics of the slow and fast traders

In the examples we consider in the next section, part of the population has an inverse memory length $r \ll 1$ and another part of the population has an inverse memory length $r = 1$. A quick look at the EWA learning update rules (see Eq. (2.2)), shows that to change its preference toward a market by a quantity of order $O(1)$, a trader with inverse memory length $r = 0.01$ will need $O(1/r) = O(100)$ rounds while a trader with finite memory will only need $O(1)$ rounds. This timescale separation between the dynamics of those two types of traders is the reason why we choose to call the traders with an inverse memory $r = 1$ *fast traders* and the traders with an inverse memory $r = 0.01$ *slow traders*. As a consequence of this timescale separation, the slow traders will be seen as static by the fast traders, which will simplify our analysis. The

market mechanisms and the learning dynamics of the traders are *exactly* the same as in Chap. 2; the fictitious play parameter α is set to 1 throughout this chapter. This choice simplifies the equations obtained from the Kramers-Moyal expansion for the dynamics of the slow traders. To study the dynamics of fast and slow traders, we will need to use two different methods. On the one hand, the natural way to describe the dynamics of slow traders is to perform a Kramers-Moyal expansion similar to the one we did in Apps. 2.C, Sec. 3.A and Sec. 5.C.1. On the other hand, the memory length of fast traders is not small $r = 1$ and this rules out using a Kramers-Moyal expansion (which relies on the smallness of r) to derive a Fokker-Planck equation for the dynamics of fast traders. Fortunately the short memory case $r = 1$ can be analysed directly because the traders' preferences become instantaneously determined by their payoffs, as we now show.

4.2.1 Update equations for fast traders

When looking at the expression of the preferences update rules in Eq. (2.2) for fast traders ($r = 1$) who use fictitious play ($\alpha = 1$), one sees that they become:

$$A_m(n+1) = \begin{cases} \mathcal{S}(n) & \text{if the trader chose market } m \text{ in round } n \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

Here $\mathcal{S}(n)$ is the score the trader earned at time n and traders use the logit rule described in Eq. (2.3) to choose the next market they will trade with. In the rest of this subsection, we will derive the probability to trade at the first market step by step. First, from the description of the trading process in Sec. 2.2 we can calculate the probability density function of the profit $\mathcal{S}$ of a trader who places an order to buy ($\tau = \mathrm{a}$) or sell ($\tau = \mathrm{b}$) at market m is:

$$\mathbb{P}(S \mid \tau, M=m) = \mathcal{V}(\tau,m)\mathcal{M}(\tau,m,f_m)\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(S-\pi_\tau)^2}{2\sigma^2})\theta(S) + \delta(S)\left(1-\mathcal{V}(\tau,m)\mathcal{M}(\tau,m,f_m)\right) \tag{4.2}$$

Therefore, for a trader that has preference to buy p_b, the probability to get a profit S when trading at market m is:

$$\mathbb{P}(S \mid M=m) = p_\mathrm{b}\mathbb{P}(S \mid \mathrm{b}, M=m) + (1-p_\mathrm{b})\mathbb{P}(S \mid \mathrm{a}, M=m) \tag{4.3}$$

In the expressions above, $\mathcal{V}(\tau,m)$ and $\mathcal{M}(\mathrm{a},m,f_m)$ are respectively the probability to send a valid order and the probability for the validated order to be matched so that a trade can take place (see Appendix 2.A). Also $\theta(\cdot)$ is the Heaviside function which is 0 if its argument is negative and 1 otherwise and $\delta(\cdot)$ is the Dirac δ-function. Let us consider $\mathbb{P}(M=m,n \mid M=m,n-1,f_m,p_\mathrm{b})$ which is "the probability that a trader with preferences to buy p_b goes to market m at time step t, knowing that he went to the same market m at time step $n-1$ and that the ratio of buyers over sellers in this market was f_m". For readability we shorten this by $\mathbb{P}(m \to m, f_m, p_\mathrm{b})$:

$$\begin{aligned}\mathbb{P}(m\to m,f_m,p_\mathrm{b}) &= \int \mathrm{d}S\,\mathbb{P}(M=m,n \mid M=m,n-1,S)\mathbb{P}(S \mid M=m,p_\mathrm{b}) \\ &= p_\mathrm{b}\int_0^\infty \mathcal{V}(\mathrm{b},m)\mathcal{M}(\mathrm{b},m,f_m)\frac{1}{\sqrt{2\pi}\sigma}\frac{\exp(-\frac{(S-\pi_b)^2}{2\sigma^2})}{1+\exp(-\beta S)} \\ &\quad + (1-p_\mathrm{b})\int_0^\infty \mathcal{V}(\mathrm{a},m)\mathcal{M}(\mathrm{a},m,f_m)\frac{1}{\sqrt{2\pi}\sigma}\frac{\exp(-\frac{(S-\pi_\mathrm{a})^2}{2\sigma^2})}{1+\exp(-\beta S)}\end{aligned} \tag{4.4}$$

Since after having traded at market m, the preferences of a trader are given by Eq. (4.1), we deduce that the probability for a trader to trade at market m knowing that s/he previously traded at market m and achieved profit S is given by $\mathbb{P}(M=m,n \mid M=m,n-1,S) = \frac{1}{1+\exp(-\beta S)}$. If we denote by $m^\dagger$ the opposite market to m (*i.e.* $m^\dagger = 1$ if $m=2$ and $m^\dagger = 2$ if $m=1$) then the probability to go from market m at time step $n-1$ to market $m^\dagger$ at time step n is:

$$\mathbb{P}(m\to m^\dagger, f_m, p_\mathrm{b}) = 1 - \mathbb{P}(m\to m, f_m, p_\mathrm{b}) \tag{4.5}$$

We then deduce from the law of large numbers that in the large population limit, $\bar{p}^{(c)}(n)$, the fraction of fast traders from class c going to market 1 at time step n can be calculated from the following recursive equation:

$$\bar{p}^{(c)}(n) = \mathbb{P}(1 \to 1, f_1, p_{\mathrm{b}}^{(c)})\bar{p}^{(c)}(n-1) + \mathbb{P}(2 \to 1, f_2, p_{\mathrm{b}}^{(c)})(1 - \bar{p}^{(c)}(n-1)) \tag{4.6}$$

This equation still involves the ratio of buyers over sellers in market m, f_m, which depends on the distribution of both fast and slow traders across both markets.

4.2.2 Distribution of the scores of the slow traders

When one performs a Kramers-Moyal expansion of the slow traders' master equation truncated at the second order, one sees that the dynamics of the distribution of difference $\Delta A = A_1 - A_2$ between the scores at market 1 and market 2 is:

$$\begin{aligned}\partial_t \mathbb{P}(\Delta A^{(c)}, t) = &-\partial_{\Delta A^{(c)}}\left[\mu^{(c)}(\Delta A^{(c)}, f_1, f_2)\mathbb{P}(\Delta A^{(c)}, t)\right] \\ &+ \frac{r}{2}\partial_{\Delta A^{(c)}}\partial_{\Delta A^{(c)}}\left[\Sigma^{(c)}(\Delta A^{(c)}, f_1, f_2)\mathbb{P}(\Delta A^{(c)}, t)\right]\end{aligned} \tag{4.7}$$

where the expressions of $\mu^{(c)}(\Delta A^{(c)}, f_1, f_2)$ and $\Sigma^{(c)}(\Delta A^{(c)}, f_1, f_2)$ are given in Appendix 4.A. Note that the dynamics of the slow and fast traders are coupled by the values of the aggregates f_1 and f_2. Moreover, it will be useful for the rest of this section to note that the stationary solution of the Fokker-Planck equation (4.7) has a closed form expression:

$$\mathbb{P}(\Delta A^{(c)}) \propto \frac{1}{\Sigma^{(c)}(\Delta A^{(c)}, f_1, f_2)} \exp\left(\frac{2}{r}\int_0^y \frac{\mu^{(c)}(y, f_1, f_2)}{\Sigma^{(c)}(y, f_1, f_2)}\mathrm{d}y\right) \tag{4.8}$$

where the sign $\propto$ means that the equation above is true up to a normalization constant. With the equation for the dynamics of both fast traders and slow traders (4.6),(4.7) it only remains to find the aggregates f_1 and f_2 to calculate the distribution of scores in the whole population. To find the value of the

aggregates, we use an iterative approach. We start with an initial guess for the aggregate from which we calculate the number of fast and slow traders in each markets with the total number of slow (resp. fast) traders being constant. We use those results to get a new estimate of the market conditions (using some damping to ensure convergence). We stop the procedure when the market conditions have converged.

4.2.3 Deterministic dynamics for the slow traders

In this chapter, to determine the behaviour of traders, we rely on the homogeneous population dynamics of slow traders. Thanks to the fact that the relaxation time of slow traders is much longer than the relaxation time of fast traders, the fluctuations in their preferences are very small. In the limit $r \to 0$ they can be neglected and we can write the following equation for the deterministic dynamics of the preferences for slow traders.

$$\dot{\Delta A}^{(c)} = \mu^{(c)} \left(\Delta A^{(c)}, f_1(\Delta A^{(1)}, \Delta A^{(2)}), f_2(\Delta A^{(1)}, \Delta A^{(2)}) \right) \tag{4.9}$$

where the aggregates $f_m(\Delta A^{(1)}, \Delta A^{(2)})$, $m \in \{1, 2\}$ are the ratios of buyers over sellers at market m. To deduce the fraction of slow traders in both markets we use the score differences, and to deduce the fraction of fast traders in both markets we use the fixed point equation (4.6). Thanks to the timescale difference between the fast traders and the slow traders, we do not need to take into account the time needed by the fast traders to relax to the fixed point of Eq. (4.6) which is confirmed by the consistency between the simulation results and our analytics shown in Fig. 4.1.

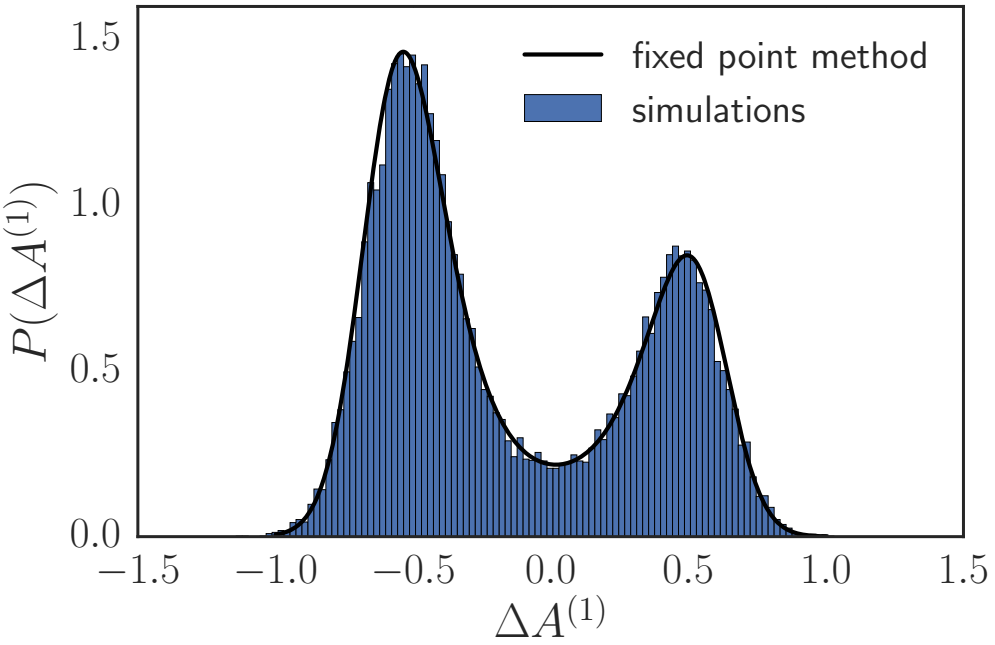

Figure 4.1: Comparison between the Fokker Planck distribution obtained with the fixed point method described in Sec. 4.2.2 and multiagent simulations. The preferences to buy of the traders in the two classes are $p_{\mathrm{b}}^{(1)} = 1 - p_{\mathrm{b}}^{(2)} = 0.2$, the market biases are $\theta_1 = 1 - \theta_2 = 0.3$, the forgetting rate is $r = 0.05$ and half of the traders are fast. The number of traders (including both fast and slow traders) is 10^4 in the multiagent simulation.

4.2.4 Onset of strong segregation

In the above setup of a system of fast and slow traders, the notion of segregation – as a separation of a population into subgroups that have different preferences over long timescales – clearly makes sense only for the slow traders. To assess the existence of weak segregation among these traders, as in Chap. 2 and 3, one just needs to count the number of peaks in the distribution of scores *i.e.* the number of fixed points of the single traders' dynamics. For strong segregation however, we need to know for each of the fixed points whether it corresponds to a peak of order 1 or to an exponentially small peak. To assess

whether the system is strongly segregated or otherwise, we need to calculate the action difference between the two (or more) fixed points of the dynamics as explained in Fig. 2.6. Luckily for a one-dimensional Langevin dynamics such as the one modelled in Eq. (4.7) the minimal action path can be calculated analytically. This means that considering the stable fixed points of the single traders' dynamics $\Delta A_1^\star$ and $\Delta A_2^\star$, the conditions that the action difference between those two paths is zero (see Sec. 2.5) is equivalent to:

$$\Delta\mathcal{A}_{1\to 2} = \int_{\Delta A_1^\star}^{\Delta A_2^\star} \frac{\mu(y, f_1, f_2)}{\Sigma(y, f_1, f_2)} \mathrm{d}y = 0 \tag{4.10}$$

where we have dropped the (c) superscript indicating the class to lighten the notation. Using the equation above, we find the strong segregation temperature by plotting the action difference between the fixed points of the single trader dynamics when the system is weakly segregated. When the quantity $\Delta\mathcal{A}_{1\to 2}$ is equal to zero, this means we are at the strong segregation threshold.

4.3 Results

4.3.1 Heterogeneity in memory

To get an intuition for the influence of memory heterogeneity on the spontaneous emergence of segregated preferences, we looked at the variations of the segregation thresholds (both strong and weak) with the fraction of traders who only remember the last trading round ($r = 1$). In Fig. 4.1 we plot the segregation thresholds for the slow traders *i.e.* $r \ll 1$. We observe that *the critical inverse intensity of choice* $1/\beta_\mathrm{c}$ *of both the strong and weak segregation thresholds decreases with the fraction of slow traders.* This is certainly due to the slow traders taking advantage of the fast traders. As a consequence, they earn more profit at their profit oriented market and will prefer to trade there (see

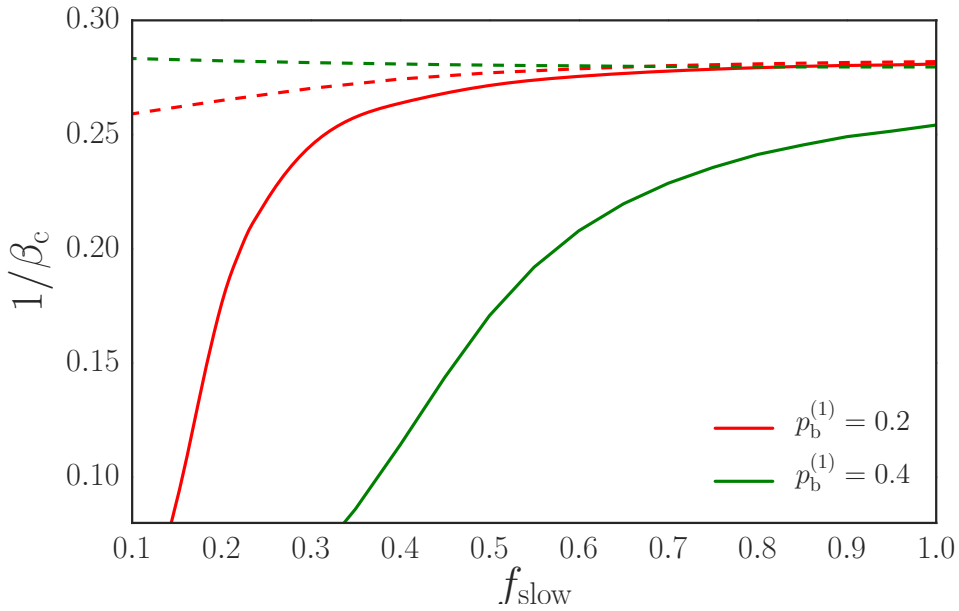

Figure 4.1: Segregation thresholds for the slow traders as a function of the fraction f of slow traders when the fast traders have the same intensity of choice as the slow traders. Market biases: $\theta_1 = 1 - \theta_2 = 0.3$ and symmetric probabilities to buy $p_b^{(1)} = 1 - p_b^{(2)}$.

Fig. 4.3). We also observe in Fig. 4.1 that when the fraction of slow traders is close to zero, the strong segregation threshold becomes small and has a strong dependence on f. Quantitatively, Fig. 4.1 shows that adding fast traders decreases the strong segregation threshold quite drastically, from $1/\beta_c \approx 0.28$ for a population consisting exclusively of slow traders, to $1/\beta_c \approx 0.15$ when the population contains only 20% of slow traders. This is not the case for weak segregation for which the segregation threshold only changes from $1/\beta_c \approx 0.28$ when the population has only slow traders to $1/\beta_c \approx 0.259$ when the population is composed of 10% of slow traders and 90% of fast traders.

The above observations can be explained by the fact that, to strongly segregate, traders have to self-organise in such a way that the buyers/sellers ratio in each market verifies self-consistency conditions similar to those in Chap. 2. But when the fraction of slow traders is decreased, the range of aggregates that

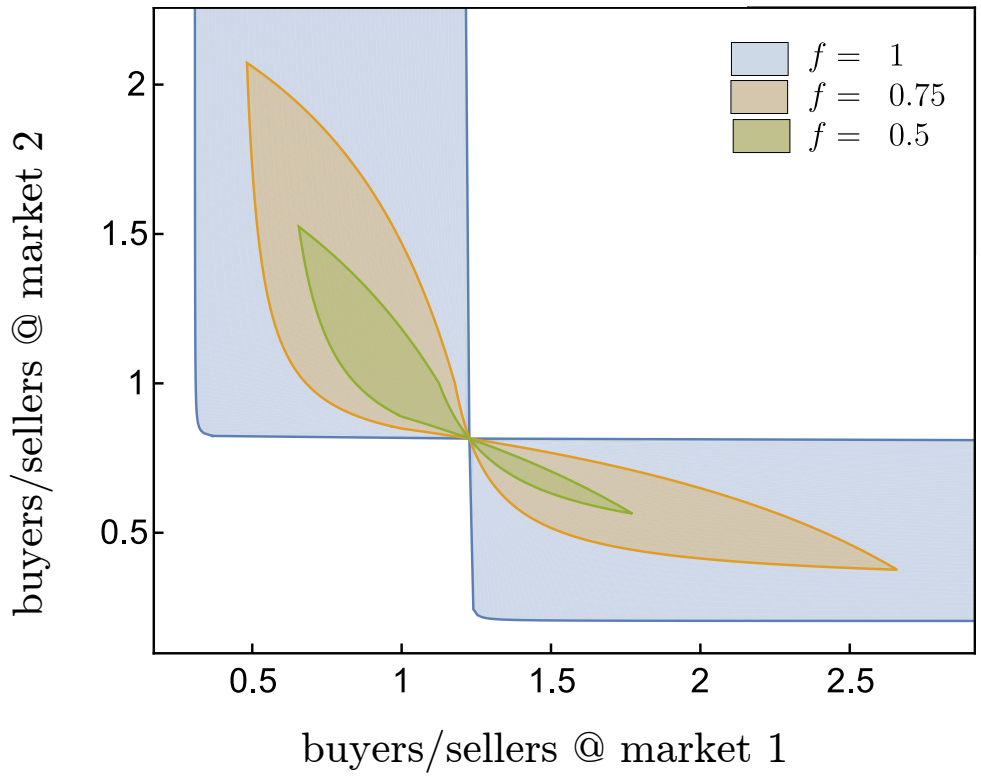

Figure 4.2: Buyers/sellers ratio which can be reached by the slow traders if they self-organise when the total population has a fraction f of slow traders, and a fraction $1 - f$ of fast traders that have an intensity of choice $\beta = 0$, *i.e.* choose market 1 exactly half of the time. As expected, as the fraction of slow traders decreases, the zone of "available" market condition gets smaller and smaller. The traders' probabilities to buy and the market biases are the same as in Fig. 4.1

can be reached gets smaller because the influence of self-organisation of the slow traders on the buyers/sellers ratio in the markets decreases. We show this in Fig. 4.2 where the possible buyer/seller ratios in each of the markets are plotted for different fractions of slow traders.

Another point of interest is how the preference distribution changes as the fraction of slow traders varies. To answer this question, we plotted in Fig. 4.3

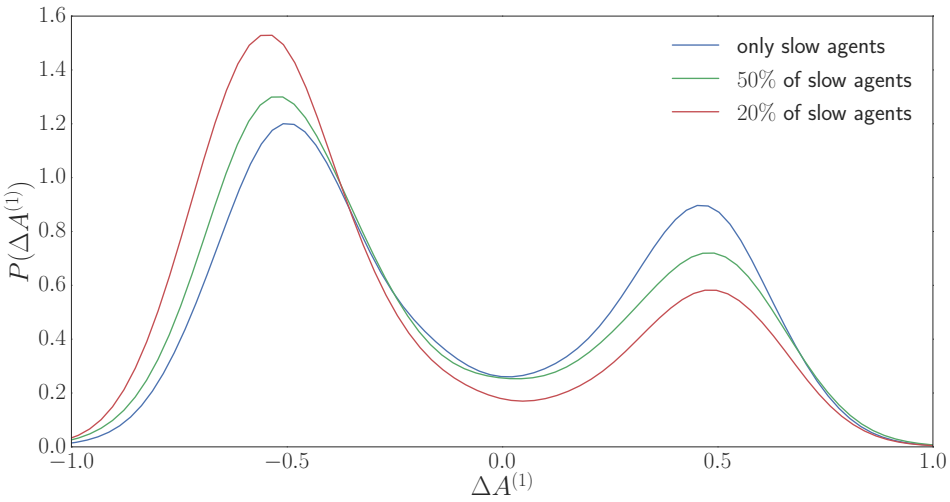

Figure 4.3: Distribution of the preferences of the slow traders for different fractions of fast traders in the population, as obtained from numerical simulations. We can see that as the fraction of fast traders increases, the number of slow traders with a preference for the profit oriented market increases. The other parameters are the same as in Fig. 4.1.

the distribution of the preferences of the slow traders in the steady state of their learning dynamics. We observe that as the fraction of slow traders decreases, *i.e.* the fraction of fast traders increases, there are more and more traders with negative preferences (they prefer to trade at the profit oriented market). This is a sign that the slow traders are taking advantage of the fast traders which enables them to trade more often at their profit oriented market.

We now look at how the thresholds shown in Fig. 4.1 vary depending on the market biases and the probability to buy of the traders $p_{\mathrm{b}}^{(1)} = 1 - p_{\mathrm{b}}^{(2)}$. In Fig. 4.4(a,b) we show the segregation thresholds for a population composed of only slow traders, a population formed of half slow and half fast traders and a population composed of 20% of slow traders and 80% of fast traders. For this range of parameters, we observe that the segregation thresholds decrease with the fraction of slow traders. We also observe in Fig. 4.4(a,b) that the effect of the variation of both the market biases and the probabilities to buy

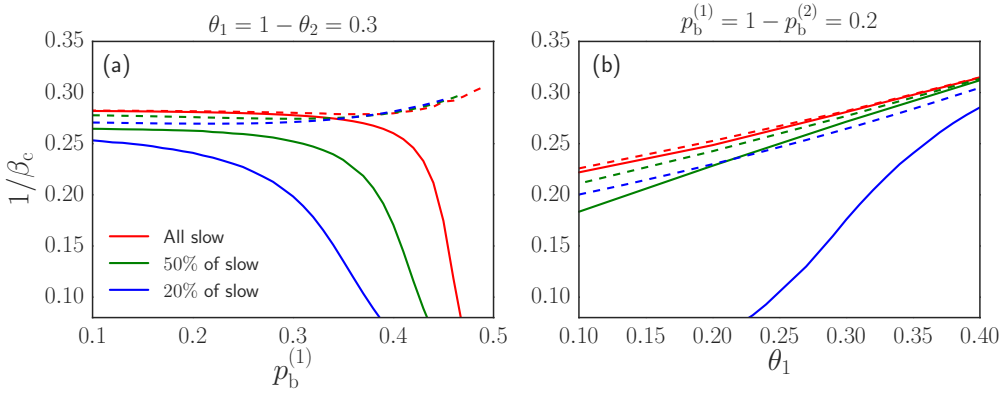

Figure 4.4: Thresholds for both strong (full line) and weak (dashed line) segregation as a function of $p_b^{(1)} = 1 - p_b^{(2)}$ (left panel) and $\theta_1 = 1 - \theta_2$ (right panel).

is larger on the strong segregation threshold than on the weak segregation threshold. This suggests that the effects described in the previous paragraph for $\theta_1 = 1 - \theta_2 = 0.3$ and $p_b^{(1)} = 1 - p_b^{(2)}$ should generalise to other values of $\theta_1 = 1 - \theta_2$ and $p_b^{(1)} = 1 - p_b^{(2)}$. Moreover, extrapolating in Fig. 4.4(b) to $\theta_1 = 0.5$ suggests – and this is confirmed by our numerics – that for two fair markets when $\theta_1 = \theta_2 = 0.5$, the fraction of fast traders does not affect the segregation thresholds. The reason behind this is the symmetry of the markets. Indeed, in such a case, slow and fast traders both go to the first and the second market with probability one half each. This keeps the ratio of buyers over sellers to 1 in each market, independently of the intensity of choice of the fast and slow traders and the fraction of fast traders. As one then moves away from the case of two fair markets, by decreasing $\theta_1 = 1 - \theta_2$, the effect of the fast traders on the segregation of the slow traders is *continuously increasing*.

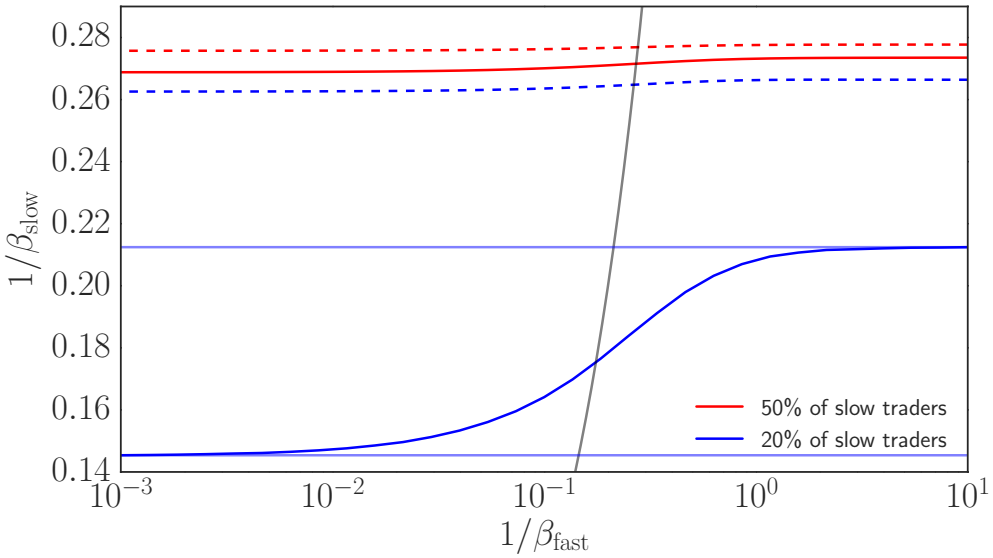

Figure 4.5: Strong (full line) and weak (dashed lines) segregation thresholds for the slow traders depending on the intensity of choice of the fast traders, β_{fast}, for classes with two different fractions of slow traders as shown. The thin lines indicate for one case the strong segregation thresholds in the two limits where the fast traders best respond to their preferences ($\beta_{\text{fast}} = \infty$) and where they choose randomly ($\beta_{\text{fast}} = 0$). In the plots, the probabilities to buy of traders are $p_{\text{b}}^{(1)} = 1 - p_{\text{b}}^{(2)} = 0.2$ and the market biases are $\theta_1 = 1 - \theta_2 = 0.3$. The grey line corresponds to $\beta_{\text{fast}} = \beta_{\text{slow}}$ and its intersection with the strong and weak segregation curves gives the segregation thresholds shown in Fig. 4.1 for $p_{\text{b}}^{(1)} = 0.2$.

4.3.2 Heterogeneity in the intensities of choice and memory length

To complete this initial analysis we investigate, in Fig. 4.5, how the segregation threshold varies when not just the memory length but also the intensity of choice of the fast traders is different from that of the slow traders. The first comment to make is that the segregation threshold increases with the inverse intensity of choice of the fast traders $1/\beta_{\text{fast}}$. We also see that while changes

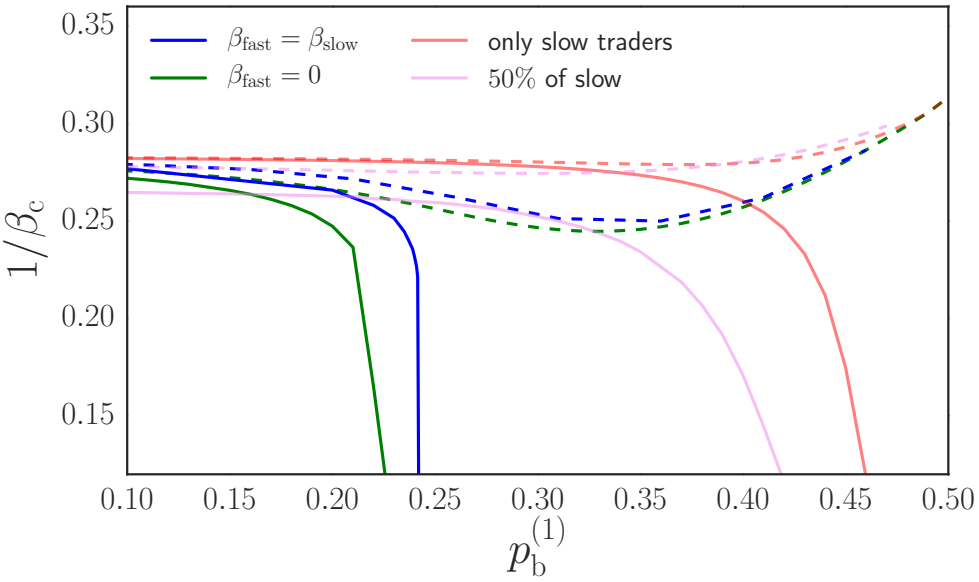

Figure 4.6: The strong (full line) and weak (dashed line) segregation thresholds when *all the traders in the first class are fast* and all the traders in the second class are slow. We consider both the case of fast and slow traders having the same intensity of choice $\beta_{\text{fast}} = \beta_{\text{slow}}$ (blue curve) and the case where the fast traders have a zero intensity of choice $\beta_{\text{fast}} = 0$ *i.e.* they choose their next market randomly. The segregation thresholds for a system with the same number of fast traders with $\beta_{\text{fast}} = \beta_{\text{slow}}$ evenly distributed between classes, and for a system with only slow traders, are shown for comparison.

in the intensity of choice of the fast traders have almost no influence on the weak segregation threshold, their influence on the strong segregation threshold is much more substantial. This is in line with our previous observations that the presence of the fast traders has more influence on the strong than on the weak segregation.

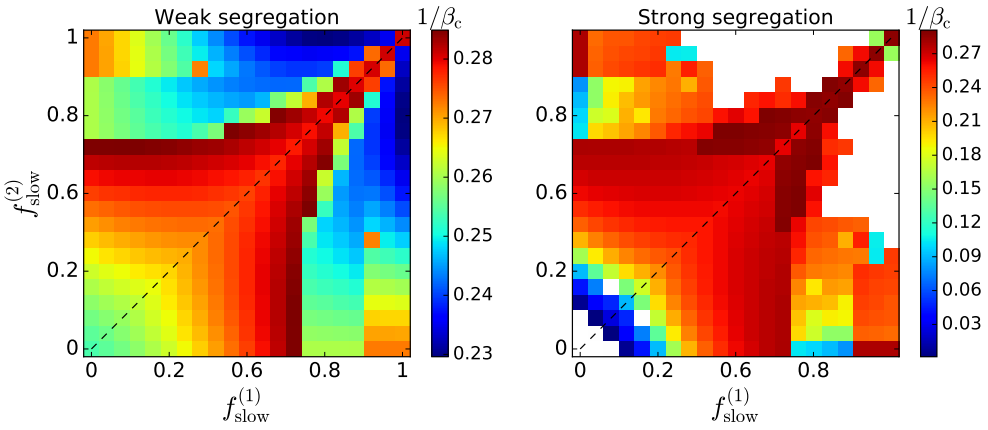

Figure 4.7: Inverse intensity of choice at which weak (left panel) and strong (right panel) segregation is observed for the first time, as a function of the fraction of slow traders in both of each class c, $f^{(c)}_{\text{slow}}$. The parameters used for this plot are $\theta_1 = 1 - \theta_2 = 0.3$ and $p^{(1)}_{\text{b}} = 1 - p^{(2)}_{\text{b}} = 0.2$. The white squares on the right indicate parameter values where we could not calculate the strong segregation temperature.

4.3.3 Heterogeneity among the classes of traders

In the previous subsection, we observed that adding the same number of fast traders in both classes decreases the segregation thresholds of the slow traders. In this subsection, we want to check if this phenomenon persists when the fractions of slow traders in class 1 and class 2 are different. We first look at the extreme case when all the traders in the first class are fast and all the traders in the second class are slow. To do so, in Fig. 4.6 we plot both the strong and weak segregation thresholds for this scenario. In order to understand the influence of the distribution of the slow traders among the classes, we also plot the segregation thresholds when there is the same overall fraction of fast traders in the population (50%) and they are evenly distributed in the two classes. The first thing we note is that when class one is composed of only fast traders and class two is composed of only slow traders, both the strong and weak segregation thresholds are lower than when there are 50% of slow traders

in each class. This suggests that an imbalance in the fraction of slow traders between the two classes impedes segregation. In fact, for the parameters we use in Fig. 4.6, once the puying probability $p_{\mathrm{b}}^{(1)}$ of class 1 is larger than ≈ 0.25, strong segregation no longer takes place at any intensity of choice of the slow traders. As is shown in Fig. 4.6 this is a phenomenon which we observe only for much higher values of $p_{\mathrm{b}}^{(1)}$ when both classes contain 50% of fast traders. To summarise the results of this discussion, not just the total fraction of fast traders but also the distribution of fast traders among the two classes has to be taken into account to understand segregation behaviour.

The findings of the previous paragraph make one wonder about the variations of the segregation thresholds with the fraction of traders in both the first and second classes. We show some results for this in Fig. 4.7. Since the market biases as well as the probability of buying of the traders are symmetric, the density plots of Fig. 4.7 are symmetric under interchange of $f_{\mathrm{slow}}^{(1)}$ and $f_{\mathrm{slow}}^{(2)}$. From those plots, we can identify two factors that influence segregation: (i) the difference between the fraction of fast traders in class 1 and in class 2, which tend to decrease the segregation thresholds and (ii) the overall fraction of fast traders in the whole population, which tends to decrease the segregation thresholds as well.

4.4 Conclusion

In this chapter, we investigated the persistence of segregation in a population where the intensity of choice, as well as the memory of the traders, are heterogeneous. After having introduced in Sec. 4.1 our motivation, which is to take into account information heterogeneity in our model, we discussed how we model such heterogeneity in Sec. 4.2. Then, we explored the consequences of various sources of heterogeneity in Sec. 4.3, proceeding in three stages. In

Sec. 4.3.1 we observed that adding the same fraction of traders with short memory in each class decreases the segregation thresholds for the traders with a long memory (see Fig. 4.5). We saw in particular that the variations with respect to the fraction of fast traders added are much larger for the strong than for the weak segregation threshold. In Sec. 4.3.2 we investigated the effects of adding heterogeneity in the intensity of choice and observed that a larger intensity of choice for fast traders decreases the segregation thresholds of slow traders. In Sec. 4.3.3 we explored how the segregation thresholds vary when the fraction of traders with short memory is different within the two classes. While showing some non-trivial structure, the results conform to the general trend that adding traders with short memory tends to lower the segregation thresholds (see Fig. 4.1). We argued that this is because the addition of fast traders reduces the range of possible market conditions that slow traders can reach by self-organising (see Fig. 4.2). The fact that the effect described above is much stronger for strong rather than for weak segregation (it can even make strong segregation disappear as observed in Fig. 4.4) confirms this intuition.

A natural continuation of this work would be to add even more sources of heterogeneity in the population of traders. For example, in the model we studied, we only considered the effect of traders with two different memory lengths while one could think of adding even more different values of the memory length. Then, it would be interesting to check if the non-monotonicity of the payoffs with respect to the information of traders observed by Toth *et al.* in [98] also takes place for a population composed of traders with several different memory lengths. It would also be appealing to see if the phenomenon of segregation persists when traders learn their favorite market with a different algorithm. Some suggestions for alternative learning dynamics can be found in [22, 40].

Another interesting extension of this work would be to add more heterogeneity in the buying probabilities of the traders. In the models presented in Chap. 2 and 3 as well as this chapter we only considered a population divided into two

classes of equal size, each of which has a specific probability to buy p_{b}. It would be interesting to see what happens when the probability to buy for each trader is fixed randomly from some nontrivial distribution at the beginning of each simulation *i.e.* quenched, and how this influences both the strong and weak segregation thresholds.

Appendix

4.A Coefficients of the Kramers-Moyal expansion

In this appendix, we give the expression of the dirft and variance which appear in the Kramer–Moyals expansion in Eq. (4.7)

$$\begin{aligned}\partial_t \mathbb{P}(\Delta A^{(c)}, t) = &-\partial_{\Delta A^{(c)}} \left[\mu^{(c)}(\Delta A^{(c)}, f_1, f_2)\mathbb{P}(\Delta A^{(c)}, t)\right] \\ &+ \frac{r}{2}\partial_{\Delta A^{(c)}}\partial_{\Delta A^{(c)}} \left[\Sigma^{(c)}(\Delta A^{(c)}, f_1, f_2)\mathbb{P}(\Delta A^{(c)}, t)\right] \end{aligned} \tag{4.11}$$

$$\mu^{(c)}(\Delta A^{(c)}, f_1, f_2) = \mathcal{P}_1^{(c)}(f_1)\sigma_\beta(\Delta A^{(c)}) - \mathcal{P}_2^{(c)}(f_2)\sigma_\beta(-\Delta A^{(c)}) - \Delta A^{(c)} \tag{4.12}$$

$$\begin{aligned}\Sigma^{(c)}(\Delta A^{(c)}, f_1, f_2) = &\left(\mathcal{Q}_1^{(c)}(f_1) - 2\Delta A^{(c)}\mathcal{P}_1^{(c)}(f_1)\right)\sigma_\beta(\Delta A^{(c)}) \\ &+ \left(\mathcal{Q}_2^{(c)}(f_2) + 2\Delta A^{(c)}\mathcal{P}_2^{(c)}(f_2)\right)\sigma_\beta(-\Delta A^{(c)}) + {\Delta A^{(c)}}^2 \end{aligned} \tag{4.13}$$

CHAPTER 5

Stochastic evolution in populations of ideas

In the previous chapters, we focused on the effects of the stochasticity of payoffs on the convergence of EWA learning in a large game. To simplify our analysis we neglected the stochasticity coming from the finite size of the populations (we assumed that the aggregates do not fluctuate). However, as pointed out in Ref. [35], finite size effects can impact learning of agents significantly. To investigate to which extent finite size effects affect the learning dynamics of agents, in this chapter we consider a simple model of players learning how to play a 2×2 normal form game using a slightly modified version of EWA learning.

5.1 Introduction

The study of games in non-cooperative game theory has traditionally focused on the analysis of their equilibrium points, in particular the celebrated Nash equilibria [69, 68] already encountered in the Chap. 2 of this thesis (see Eqns. (2.7a, 2.7b)). These are the points in strategy space that fully ratio-

nal players choose, based on full information of the game and assuming that their opponents act fully rationally as well. At a Nash point no player can increase their payoff by *unilaterally* changing their strategy. These ideas provide a natural first approach to the analysis of games, and they are mathematically convenient as they do not involve any actual dynamics. On the other hand the scope of such equilibrium concepts is naturally limited. Indeed, in the former chapters, we observed that the dynamics of players learning to play a simple game could be captured by the Nash equilibrium only in some specific cases (see Sec. 2.4.1). This is confirmed by behavioural economics experiments which show that real-world players do not behave fully rationally in repeated games, and suggest that inductive learning from past experience may be a better model than the assumption of full rationality [19, 18].

In many models of dynamic learning, players do not find the mutually optimal strategy immediately; in fact they potentially never do. Instead they initially try out the different actions available to them, and attempt to learn from past experience. Players assess the success or otherwise of individual strategies and then choose those that worked well in the past. Their opponents adapt as well, and strategies that may have performed well previously can become less successful when the opponents' propensities have changed. This generates a coupled dynamics between the players, and it is not clear a-priori if and when such dynamics converge to Nash points. Indeed, work on games of low and high complexity has suggested that learning may result in chaotic motion [42, 93, 13, 87], in some cases with very high dimensional attractors. Situations in which systems of this type settle down to unique well-defined fixed points then seem to be the exception rather than the rule.

Learning and adaptation based on past experience can be interpreted as an evolutionary process of 'ideas' in the minds of the players. Börgers and Sarin, for example, write [10] *'Decision makers are usually not completely committed to just one set of ideas [...]. Rather [...] several possible ways of behaving are*

present in their minds simultaneously. Which of these predominate, and which are given less attention, depends on the experiences of the individual. The change which the "population of ideas" in the decision maker's mind undergoes may be analogous to biological evolution.' Similar approaches have also been used in models of language evolution; see e.g. Blythe et al [9]. In the context of a game the evolutionary process in a population of ideas broadly works as follows: each player carries in his or her mind a mixed populations of ideas. These represent the different actions (pure strategies) he or she can take in the game. Different ideas will be present in the player's mind in different proportions. At each instance of the game each player pulls out one idea (action) out of their mind at random, and uses it in the game. The ideas that are more frequent in the player's mind will be used more often than those which are present less in the population. The composition of the player's mind thus represents their mixed strategy. Over time the player learns from past experience, and the population of ideas in their mind undergoes an evolutionary process: less successful ideas are displaced by more successful strategies. This is illustrated in Fig. 5.1, and akin to well-known birth-death processes in evolutionary dynamics [100]. It is hence no surprise that the equations governing multi-player learning can be very similar to those used to model evolutionary dynamics [10, 88].

Most existing analogies between learning and evolutionary dynamics are at the level of deterministic differential equations though, formally describing the dynamics of infinite populations. At the same time, evolutionary dynamics in finite populations shows several phenomena that arise solely from intrinsic stochasticity. These effects include noise-driven fixation and extinction, which are not captured by deterministic approaches. A substantial amount of work is available on the dynamics of stochastic birth-death processes, including an analytical formalism to compute fixation probabilities and the times to fixation, see for example [75, 4, 3, 100].

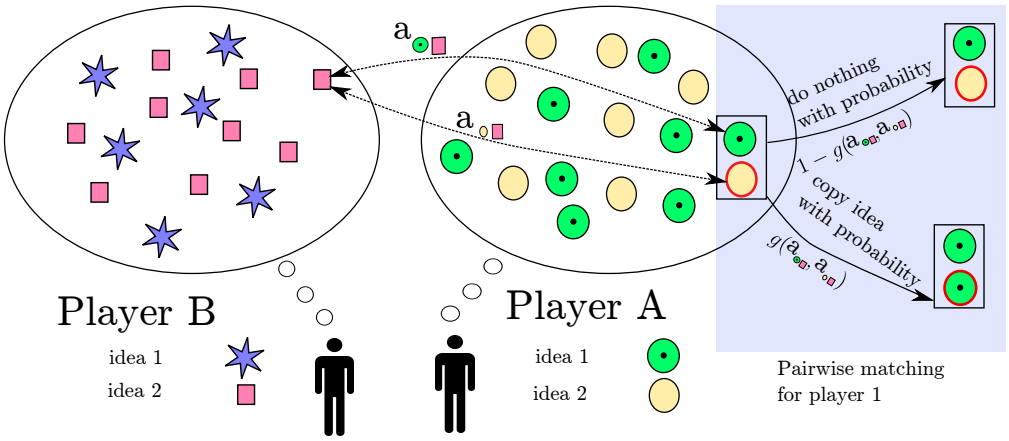

Figure 5.1: Illustration of the evolutionary process that occurs in a population of ideas: two ideas in the mind of player A are selected ($\odot$ and $\bigcirc$) as indicated by the rectangle on the right. Both ideas play against the same randomly chosen adversary idea (here $\square$) in the population of ideas of player B and the relevant payoffs are recorded, here denoted $a_{\odot\square}$ and $a_{\bigcirc\square}$. Idea $\bigcirc$ is switched to $\odot$ with probability $g(a_{\odot\square}, a_{\bigcirc\square})$ depending on these payoffs. An analogous process occurs in the population of ideas of player B. The non-negative function $g(\cdot,\cdot)$ is increasing in the first argument, and decreasing in the second. It defines the mechanics of the evolutionary process. See also the text in Secs. 5.3.3 and 5.4.1 for further details.

The main purpose of the present chapter is to develop a microscopic representation of reinforcement learning as a stochastic evolutionary process in a finite population of ideas. Ideas in this description are members of a finite populations, and undergo a birth-death process. This approach allows us to establish the analogy between learning and evolution at the level of *stochastic* population dynamics. More specifically we will define the transition rates of a birth-death process in a population of ideas, such that the deterministic description in the limit of infinite populations reproduces the so-called Sato-Crutchfield differential equations [88, 86]. We show that the notion of reproductive fitness needs to be augmented by an entropic restoring force to capture weak decision prefer-

Notation in this chapter	Notation in Chaps. 2, 3 and 4
x	mixed strategy p
i	market chosen τ
Q	attraction A
π	average score $\bar{S}$
A, B	class label c
α	forgetting rate r
$\Gamma/\alpha = \lambda^{-1}$	intensity of choice β

Table 5.1: Comparison of the notation of this chapter with the notation of Chaps. 2, 3 and 4. See Sec. 5.2.2 for an exhaustive discussion.

ences and/or memory loss in game learning. These restoring forces play a role similar to that of mutation in evolutionary dynamics. Crucially, however, the birth-death dynamics in finite populations of ideas has absorbing states so that ideas can go extinct or reach fixation. This marks a key difference compared to mutation-selection dynamics, where there are no absorbing states.

The remainder of the chapter is organized as follows. In Sec. 5.2 we briefly summarize the mathematics of the standard replicator dynamics and of the reinforcement learning dynamics we use as a basis for the evolution of ideas. In Sec. 5.3 we then introduce the birth-death process for finite populations of ideas, and we study its properties for simple symmetric games. In Sec. 5.4 we extend the analysis to two-player learning in asymmetric games. Finally in Sec. 5.5 we collect our conclusions and present an outlook towards future work. Further technical details of our analysis can be found in the Appendices. Note that in order to study the finite size effects described above in full generality, we decided to use notational conventions that differ from the previous chapters. We describe the new notation in more detail in Sec. 5.2.2 and summarise it in Tab. 5.1.

5.2 Deterministic evolutionary dynamics and adaptive learning

5.2.1 Evolutionary dynamics and replicator equations

5.2.1.1 Single-population replicator equations

The evolutionary dynamics of interacting individuals in infinite populations is frequently described by replicator [97] or replicator-mutator equations. These are deterministic ordinary differential equations. We focus on a population of individuals of S different types, $i = 1, \ldots, S$, and write $x_i(t)$ for the fraction of individuals of type i in the population at time t, and $\mathbf{x} = (x_1, \ldots, x_S)$. At all times $\sum_i x_i(t) = 1$. We assume that individuals interact in a symmetric two-player normal form game [50]. This is specified by a payoff matrix $\mathbf{A} = (a_{ij})$. The entry a_{ij} is the payoff to an individual of type i in an interaction with an individual of type j. The setup of a symmetric game is not to be confused with a game for which the payoff matrix is symmetric, i.e. its own transpose.

The average payoff per game to an individual of type i in a population of composition $\mathbf{x}$ is given by $\pi_i(\mathbf{x}) = \sum_j a_{ij} x_j$. In order to keep the notation compact, we will omit the argument $\mathbf{x}$ in the following. The standard replicator equations are then given by [50]

$$\dot{x}_i = x_i(\pi_i - \pi), \tag{5.1}$$

with $\pi = \sum_j x_j \pi_j$. These dynamics can be derived from a birth-death process in the limit of an infinite population. This will be discussed in more detail below.

5.2.1.2 Two-population replicator dynamics

The case of asymmetric games refers to situations in which different individuals take on different roles, e.g. male and female in Dawkin's battle of the sexes [34], or buyers and sellers in a stock market. In this case individuals belonging to different populations. In two-population replicator systems the fitness of individuals in population A is determined by their interaction with individuals in population B, and vice versa. Selection and evolution then occur within each population; see [50] for details. This leads to the following two-population replicator dynamics:

$$\dot{x}_i^A = x_i^A(\pi_i^A - \pi^A), \tag{5.2a}$$

$$\dot{x}_i^B = x_i^B(\pi_i^B - \pi^B), \tag{5.2b}$$

where x_i^A is the frequency with which individuals of type i occur in population A, and x_i^B the frequency with which the i-th type occurs in population B. It is important to note that the label i in either population is a simple numbering of pure strategies, *e.g.* in Dawkin's battle of the sexes $i = 1, 2$ in the populations of males may refer to 'faithful' and 'philanderer', and in the population of females the same labels may refer to 'coy' and 'fast' [34, 99].

In the above equations we have used the shorthands,

$$\pi_i^A = \sum_j a_{ij} x_j^B, \tag{5.3a}$$

$$\pi_i^B = \sum_j b_{ij} x_j^A, \tag{5.3b}$$

as well as $\pi^A = \sum_i \pi_i^A x_i^A$ and similarly $\pi^B = \sum_i \pi_i^B x_i^B$.

5.2.2 Discrete-time Sato-Crutchfield learning

Following [86, 88] we consider two players, labelled A and B repeatedly playing an asymmetric game with payoff matrices $\mathbf{A}$ ($\mathbf{B}$) for player A (B). For simplicity, we will assume that both players have the same number S of actions available, but the extension to the more general case is straightforward [86]. Hence $\mathbf{A}$ and $\mathbf{B}$ will be $S \times S$ matrices, with entries denoted a_{ij} and b_{ji}, $i, j = 1, \ldots, S$. As implied in (5.3a) and (5.3b) above, a_{ij} is the payoff to player A if she chooses action i while player B plays action j; b_{ji} is the payoff to player B in this situation.

At each instance of the game, each player $\mu \in \{A, B\}$ will choose one action. In order to monitor the relative success of the different actions, each player holds an 'attraction' for each action. We will write $Q_i^\mu(t)$ for the attraction player μ has for action i at time t. Sato-Crutchfield learning assumes a soft-max (or logit) rule to convert a set of attractions $Q_1^\mu, \ldots, Q_S^\mu$ into a mixed strategy,

$$x_i^\mu = \frac{\exp(\Gamma Q_i^\mu)}{\sum_j \exp(\Gamma Q_j^\mu)}. \tag{5.4}$$

The parameter $\Gamma \geq 0$ represents the intensity of choice as in [42, 49, 19]. When $\Gamma = 0$ attractions play no role and players choose their actions with equal probability. In the limit $\Gamma \to \infty$ players play a pure strategy that always chooses the action with the highest attraction.

In [86, 88] the preferences for the different actions are updated in discrete time. It is also assumed that a large (formally infinite) number of rounds of the game is played in between such updates, and that player A observes player B's actions and vice versa. Each agent then has full knowledge of the other agent's mixed strategy. This is a simplification of the model, which was made for convenience in [86] and results in a full deterministic dynamics. The learning dynamics remains stochastic if the number of observations made between updates is finite [41, 82].

Proceeding on the basis of a deterministic dynamics, Sato-Crutchfield learning takes the form

$$Q_i^A(t+1) = (1-\alpha)Q_i^A(t) + \sum_j a_{ij}x_j^B(t), \tag{5.5}$$

$$Q_i^B(t+1) = (1-\alpha)Q_i^B(t) + \sum_j b_{ij}x_j^A(t). \tag{5.6}$$

The parameter α describes geometric discounting over time. For $\alpha = 0$ the players have full memory of the past, and the attraction $Q_i^\mu(t)$ represents the total payoff player $\mu \in \{A, B\}$ would have achieved up to time t given the other player's actions, and if μ had always used action i. For positive values of α more recent rounds contribute more to the attraction than iterations of the game in the distant past. The parameter α is restricted to the range $0 \leq \alpha \leq 1$.

The learning rule in Eq. (5.6) is a form of EWA learning as used in previous chapters, but with two differences. Firstly, the payoffs on the r.h.s. are the expected payoffs – given the strategies used by the players – so the learning rule implicitly assumes that these are known without stochasticity and for all actions i. In the market choice game, on the other hand, the learning dynamics was driven by stochastic payoffs and these were known only for the currently chosen action. We will see below how in the population dynamics view the deterministic payoffs arise from averaging over the stochastic payoffs received in games among randomly chosen members of the population. Similarly, this averaging guarantees that the payoffs are known for all actions (except if there are no individuals at all that play a certain action, in which case this action is extinct and its payoff no longer relevant).

The second difference is in the parameterization of the learning rule in Eq. (5.6), for which we followed standard notation for Sato-Crutchfield learning here. To see the relation to the form of EWA learning studied in previous chapters (see e.g. Eqns. (2.2, 2.3)), consider Eq. (5.6). We relabel $i \to m$ (action = market chosen), $t \to n$ (discrete time = trading rounds) and call

the expected payoff on the r.h.s. $\mathcal{P}_m$ as in earlier chapters. If we then also identify α with the parameter r used previously and rescale the attractions as $A_m = \alpha Q_i^A = r Q_i^A$, leaving the agent index on A_m implicit, Eq. (5.6) becomes after multiplication by r

$$A_m(n+1) = (1-r)A_m(n) + r\mathcal{P}_m \tag{5.7}$$

This is identical to Eq. (2.2) except for the two changes discussed above, i.e. replacing stochastic payoffs by expected payoffs and assuming known payoffs for all actions. We conclude that α here plays the role of the learning rate r used in previous chapters.

To see the analogous correspondence for Γ, note that from the rescaling between A_m and Q_i^A, the link in Eq. (5.4) from attractions to strategies is governed by $\exp(\Gamma Q_i^A) = \exp((\Gamma/\alpha)A_m)$. This is identical to the exponentials in Eq. (2.3) if one identifies

$$\beta = \Gamma/\alpha \tag{5.8}$$

Note that the inverse of this parameter is defined as λ below. Consistent with the intuition that β corresponds to an inverse temperature, $\lambda = \beta^{-1}$ is a temperature-like parameter of Sato-Crutchfield learning that multiplies the entropic contributions to the learning dynamics.

5.2.3 Continuous-time limit and modified replicator equations

Combining Eqs. (5.4, 5.5, 5.6) one finds

$$x_i^\mu(t+1) = \frac{[x_i^\mu(t)]^{1-\alpha}\exp\left(\Gamma\pi_i^\mu\right)}{\sum_j [x_j^\mu(t)]^{1-\alpha}\exp\left(\Gamma\pi_j^\mu\right)}. \tag{5.9}$$

In order to derive a continuous-time limit we formally rescale the time step of learning to be Δt (so that $t+1$ on the LHS of Eq. (5.9) becomes $t+\Delta t$).

We also rescale the model parameters and write $\alpha\Delta t$ instead of α, and $\Gamma\Delta t$ instead of Γ. Then taking the limit $\Delta t \to 0$ we find

$$\dot{x}_i^\mu = \Gamma x_i^\mu \left[\left(\pi_i^\mu - \sum_j \pi_j^\mu x_j^\mu \right) - \lambda \left(\ln x_i^\mu - \sum_j x_j^\mu \ln x_j^\mu \right) \right], \tag{5.10}$$

where $\lambda = \alpha/\Gamma$. The first term on the right-hand side is the expression known from the standard multi-population replicator dynamics in Eq. (5.2a) and (5.2b). The term proportional to λ exerts a force towards a uniformly mixed strategy, $x_i^\mu = 1/S$. This 'entropic' force will be strong when either the intensity of choice is low (players tend to choose their actions at random), or when memory loss is quick (propensities do not become sufficiently different to discriminate effectively between actions).

We conclude this section by two brief, but consequential observations. First, the flow of the replicator Eqs. (5.2a) and (5.2b) can be towards stable fixed points at which one or several of the actions are not played (i.e. $x_i^\mu = 0$). This cannot occur in the Sato-Crutchfield equations when $\lambda > 0$. Any attracting fixed points must be in the interior of strategy space. Secondly we note that the Sato-Crutchfield equations (5.10) can be written in the form of conventional replicator equations

$$\dot{x}_i^\mu = \Gamma x_i^\mu \left(f_i^\mu - \sum_j x_j^\mu f_j^\mu \right) \tag{5.11}$$

by introducing a modified fitness as

$$f_i^\mu = \pi_i^\mu - \lambda \ln x_i^\mu. \tag{5.12}$$

This will be the starting point for our construction of an individual-based model for the evolution of a population of ideas.

5.3 Stochastic dynamics in finite populations: the case of symmetric games

5.3.1 Birth-death dynamics

To briefly recall the main features of simple birth death processes [100, 75] we consider a population of N individuals, each of which can be of one of two types, $i = 1, 2$. We write n for the number of individuals of type 1; the remaining $N - n$ individuals are of type 2. Evolution proceeds in this population via a continuous-time Markov process with transition rates T_n^+ from state n to state $n+1$, and T_n^- from state n to state $n-1$. In the context of evolutionary games these rates are of the general form (see for example [7])

$$T_n^+ = \frac{n(N-n)}{N} g(\pi_1, \pi_2), \tag{5.13a}$$

$$T_n^- = \frac{n(N-n)}{N} g(\pi_2, \pi_1), \tag{5.13b}$$

where $\pi_1 = [a_{11}n + a_{12}(N-n)]/N$ is the fitness of an individual of type 1 in the population, with an analogous expression for π_2. The rates scale linearly with the population size N – this is a standard choice [100, 7], which implies that time is effectively measured in units of generations. From these rates a deterministic dynamics is obtained in the limit $N \to \infty$ [100]. For large (formally infinite) populations and writing $x = n/N$, one finds

$$\dot{x} = x(1-x)\left[g(\pi_1, \pi_2) - g(\pi_2, \pi_1)\right]. \tag{5.14}$$

A commonly used choice for the function $g(\cdot, \cdot)$ is the so-called linear pairwise comparison process [7, 100],

$$g(\pi_1, \pi_2) = \frac{1}{2}\left[1 + \Gamma(\pi_1 - \pi_2)\right] \tag{5.15}$$

where the parameter $\Gamma \geq 0$ is chosen small enough to ensure that $g \geq 0$ for all x. The duplicate use of Γ is intentional, as will become clear shortly. With

the above choice of g one obtains

$$\dot{x} = \Gamma x(1-x)\left(\pi_1 - \pi_2\right), \tag{5.16}$$

Modulo the constant pre-factor Γ this is easily shown to be the replicator equation (5.1) with $S = 2$.

5.3.2 Interpretation of fitness in the linear pairwise comparison process

We digress briefly in this subsection to discuss how individuals in the above birth-death dynamics have access to their fitness, i.e. their average payoff.

A common interpretation of fitness functions of the type $\pi_i = \sum_j a_{ij}x_j$ requires a fast interaction time scale on which individuals face each other in the game [41, 82, 83]. The evolutionary dynamics is assumed to be a (much) slower process; it can therefore draw on knowledge of π_i as defined above.

One particular advantage of the linear pairwise comparison process (5.15) is that it does not require such a separation of time scales between interaction and evolution. Instead one can construct the evolutionary process as follows: for any (potential) birth-death event an ordered triplet of individuals from the population is picked (with replacement). We refer to the individuals in this triplet as "primary", "secondary" and "adversary", and denote their types by i_1, i_2, i_a. Once a triplet has been picked, the primary and secondary individual both play against the adversary and receive payoffs $a_{i_1 i_a}$ and $a_{i_2 i_a}$, respectively. The secondary individual (i_2) is then replaced by an individual of the primary type (i_1) – a combined death-birth event – with probability $g(a_{i_1 i_a}, a_{i_2, i_a})$; otherwise the system is left unchanged. For the choice of g as in Eq. (5.15) the Markov chain governing this process is then that described by the rates in Eq. (5.13a, 5.13b). This is easily demonstrated for $S = 2$. With

appropriate scaling of the rates with N we find

$$T_n^+ = \frac{n_1 n_2}{N^2}[n_1 g(a_{11}, a_{21}) + n_2 g(a_{12}, a_{22})]. \tag{5.17}$$

The term in square brackets effectively averages over the choice of adversary. Using the specific form of the linear pairwise comparison process in Eq. (5.15), this can be written as

$$T_n^+ = \frac{n_1 n_2}{N} g\left(\frac{n_1}{N}a_{11} + \frac{n_2}{N}a_{12}, \frac{n_1}{N}a_{21} + \frac{n_2}{N}a_{22}\right), \tag{5.18}$$

which demonstrates the equivalence.

5.3.3 Birth-death dynamics in a finite population of ideas

We now construct an individual-based representation of Sato-Crutchfield dynamics. Motivated by Eq. (5.12) we introduce the modified fitness

$$f_i = \sum_j a_{ij}\frac{n_j}{N} - \lambda \ln \frac{n_i}{N}, \tag{5.19}$$

which can be seen as 'entropically' penalizing ideas that occur very frequently, and favouring rarer types. Focusing on the simplest case $S = 2$ we use birth-death rates

$$T_n^+ = \frac{n(N-n)}{N} g(f_1, f_2), \quad T_n^- = \frac{n(N-n)}{N} g(f_2, f_1). \tag{5.20}$$

with g as defined in Eq. (5.15). This is a representation of Sato-Crutchfield learning in the sense that it leads to the dynamics

$$\dot{x} = \Gamma x(f_1 - f) = \Gamma x(\pi_1 - \pi) - \Gamma\lambda x\left(\ln x + s\right), \tag{5.21}$$

in the limit of infinite populations. We have written $s = -\left[x \ln x + (1-x)\ln(1-x)\right]$, and $f = xf_1 + (1-x)f_2$. Our main focus from now on will be the behaviour of this birth-death process in *finite* populations.

The parameters Γ and λ need to be chosen such that all transition rates $T_n^{\pm}$ are non-negative. Written out explicitly the transition rates in Eq. (5.20), with the definition (5.15) are

$$T_n^{\pm} = \frac{n(N-n)}{N}\frac{1}{2}\left[1 \pm \Gamma\left(\Delta\pi - \lambda \ln\frac{n}{N-n}\right)\right]. \tag{5.22}$$

where $\Delta\pi = \pi_1 - \pi_2$. Thus, we require

$$\Gamma\left|\Delta\pi - \lambda \ln\frac{n}{N-n}\right| \leq 1 \tag{5.23}$$

for all $n = 1, \ldots, N-1$. At fixed Γ, this imposes a constraint $\lambda < \lambda_c$, where $\lambda_c = \mathcal{O}(1/\ln N)$ is weakly dependent on population size; see the appendices for details. Alternatively, one could choose a manifestly positive function $g(\cdot,\cdot)$, such as $g(f_1, f_2) = [1+\exp(-2\Gamma(f_1 - f_2))]^{-1}$. The resulting dynamics is known as the Fermi process [7, 3]. While the fixed points of the resulting deterministic dynamics are the same as for the linear comparison process, the dynamics themselves are quantitatively different from Sato-Crutchfield dynamics. We therefore do not pursue this route.

The expressions in Eq. (5.22) imply $T_n^+ = T_N^- = 0$, keeping in mind that $\lim_{n\to 0} n \ln n = 0$. The states $n = 0$ and $n = N$ are therefore absorbing. Accordingly, the birth-death dynamics in the population of ideas shows fluctuation-induced extinction of ideas (or equivalently fixation). In the remainder of this section we study these fixation phenomena in the context of simple 2×2 games.

5.3.4 Application to symmetric two-player two-strategy games

We focus on three common types of games that cover the qualitatively distinct deterministic flow patterns available under replicator dynamics. The corresponding payoff matrices are given in Fig. 5.1, along with illustrations of the

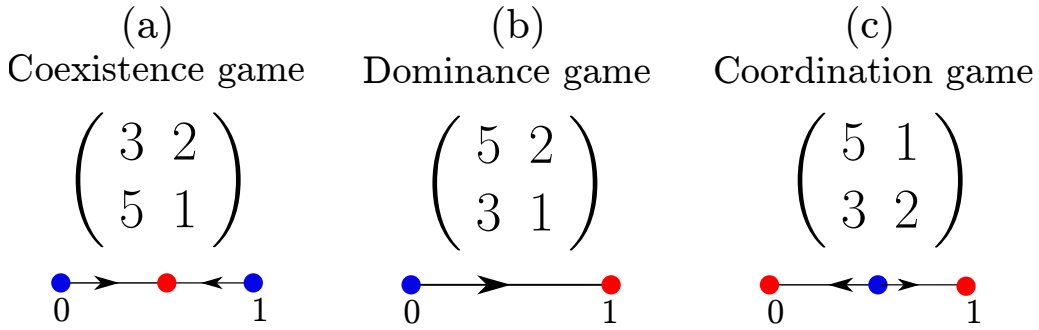

Figure 5.1: Payoff matrices **A** of the three main types of two-strategy two-player symmetric games, and their flow diagrams in $x \in [0, 1]$ under replicator dynamics.

respective replicator flow ($\lambda = 0$). The points $x = 0$ and $x = 1$ are fixed points for all games for all values of λ.

Note that it is only asymmetric games as defined in Sec. 5.2.1.2 for which the deterministic dynamics has a natural interpretation in terms of Sato-Crutchfield learning for a two-player game. Our study of symmetric games, where the only notion of game play is in the pairwise interaction of the *individuals* in a population – rather than between two distinct populations representing players in the sense of Sato-Crutchfield – is primarily a warm-up. It will help us identify some important mechanisms of the fixation dynamics, such as deterministic relaxation and activation, that will be helpful in our analysis of asymmetric games in Sec. 5.4.

Co-existence games. The boundary fixed points ($x = 0, x = 1$) are unstable for co-existence games under replicator flow, and there is a stable interior fixed point $x^\star$ where both types of ideas coexist. The memory-loss term in the Sato-Crutchfield equation ($\lambda > 0$) does not change the qualitative features of the flow; its main effect is to move the stable fixed point closer the centre of the state space, as shown in Fig. 5.2(a). For very quick memory loss ($\lambda \gg 1$)

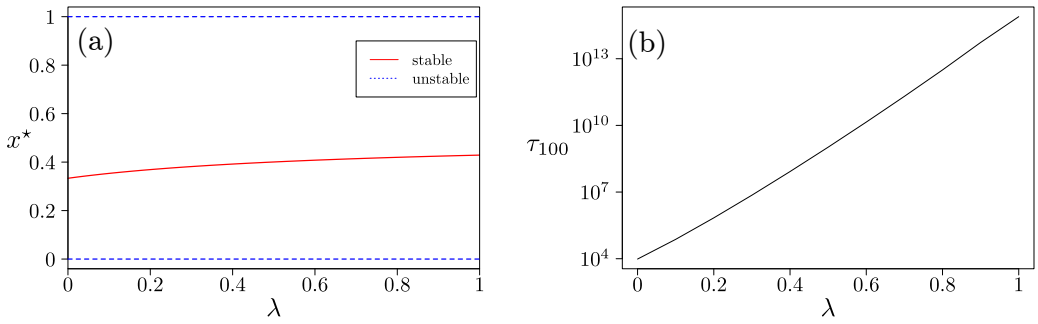

Figure 5.2: Co-existence game: (a) Location of fixed points of single-population Sato-Crutchfield learning, Eq. (5.21). (b) Mean fixation time as a function of λ in a finite population of size $N = 200$, starting at initial condition $n = 100$. The line is obtained using the known closed-form solution for simple birth-death processes, see e.g. [100]. Intensity of choice is $\Gamma = 0.1$.

the fitness f_i of either type of individual is entirely dominated by the entropic term, and both types of individuals are present with equal frequency.

The path to fixation in finite population coexistence games consists of two parts: (i) an initial relaxation to the vicinity of the interior fixed point; (ii) activation to one of the two absorbing states, driven by fluctuations; see also the appendices for further discussion. Eyring-Kramers theory [46, 37, 58] indicates that the typical time required for such an activation event grows exponentially with the height of the relevant activation barrier, and with the inverse variance of the noise, N. The height of the activation barrier is affected by the restoring force of the entropic term. Accordingly, the fixation time shown in Fig. 5.2(b) shows a strong dependence of fixation times on the model parameter λ at fixed N. The functional form is approximately exponential, suggesting a linear increase in the activation barrier with λ. This is intuitively plausible in the limit of large λ: the entropic term will dominate the dynamics, and it is linear in λ.

Dominance games. In this type of game one idea is dominant, and always has a higher payoff than the other type of idea. The replicator flow has constant sign; for the choice of payoff matrix in Fig. 5.1(b) it has an unstable fixed point at $x = 0$, and a stable fixed point at $x = 1$. The Sato-Crutchfield dynamics at $\lambda > 0$ has an additional stable interior fixed point $x^\star$, which approaches unity as $\lambda \to 0$, see Fig. 5.3(a). In finite populations the dynamics is similar to that of the coexistence game when $\lambda > 0$. After an initial relaxation towards the interior fixed point, noise drives the system to fixation. Given that the fixed point is located close to $x = 1$ for small and moderate λ, fixation will mostly occur at the upper absorbing boundary. As before fixation times increase with λ but are rather shorter than in the coexistence game, see Fig. 5.3(b). Exponential dependence of the fixation time on λ is only seen when λ is sufficiently large so that the internal fixed point is well separated from the absorbing states, or when the population size is large enough for the activation barrier to show. For small and moderate values of λ the activation barrier is too shallow relative to the noise strength for Eyring-Kramers theory to apply.

Coordination games. In addition to the trivial fixed points at the boundaries, the replicator dynamics of the coordination game has an unstable interior fixed point $x_0^\star$. With memory-loss ($\lambda > 0$) the dynamics develops a more intricate structure, see Fig. 5.4. At small but non-zero λ there are five fixed points. As λ is increased, two of these fixed points merge in a saddle-node bifurcation; we denote the corresponding value of λ by λ_c. For stronger memory loss there are three fixed points, but with reversed stability compared to the situation at $\lambda = 0$: unstable fixed points at $x = 0$ and $x = 1$, and a stable interior fixed point whose location depends on λ.

For $\lambda < \lambda_c$ and initial conditions $n/N = x > x_0^\star$, i.e. above the unstable fixed point in the lower left of Fig. 5.4(a), fixation takes place as in the dominance

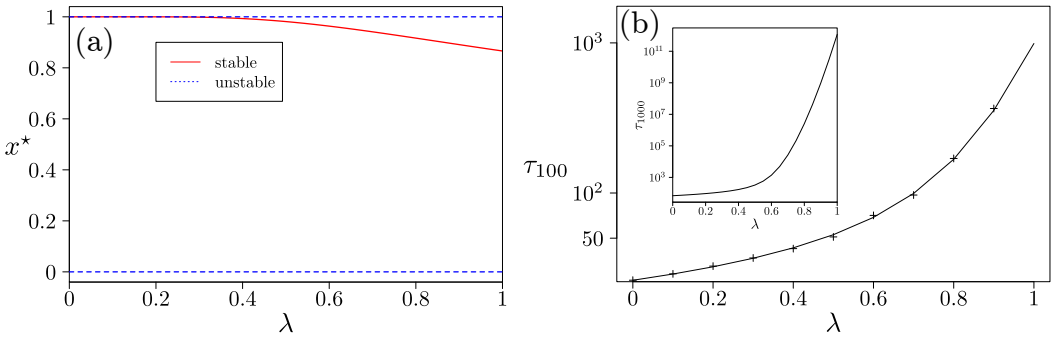

Figure 5.3: Dominance game. (a) Location of fixed points of single-population Sato-Crutchfield learning, Eq. (5.21). (b) Mean fixation time as a function of λ in a finite population of size $N = 200$, starting at initial condition $n = 100$, comparing theory (continuous line) to direct numerical simulations of the dynamics (markers) using the Gillespie algorithm. Intensity of choice is $\Gamma = 0.1$. In the inset of panel (b) we show the mean fixation time starting from $n = 1000$ for a population of size 2000, where the crossover to an exponential dependence on λ is visible at large λ.

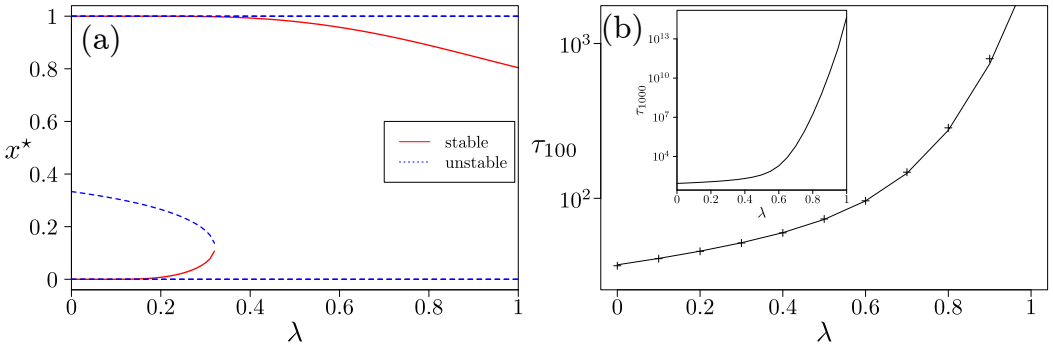

Figure 5.4: Coordination game. (a) Location of fixed points of single-population Sato-Crutchfield learning. (b) Mean fixation time as a function of λ in a finite population of size $N = 200$, starting at initial condition $n = 100$, comparing theory (continuous line) to numerical simulations of the dynamics (markers). Intensity of choice is $\Gamma = 0.1$. In the inset of panel (b) we show the mean fixation time starting from $n = 1000$ for a population of size 2000.

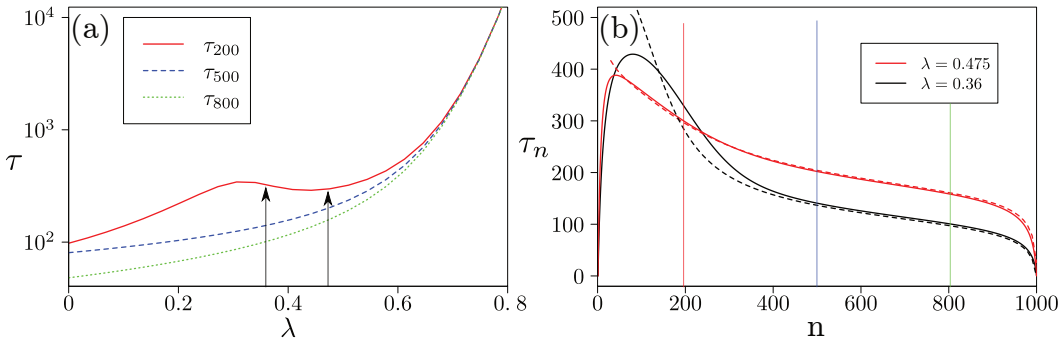

Figure 5.5: Coordination game. (a) Mean fixation time in a finite population (with $\Gamma = 0.1$) as a function of λ, the memory-loss parameter, and for different initial conditions n. (b) Mean fixation time as a function of the initial condition for two fixed values λ indicated by arrows in (a). We show data for a larger population size $N = 1000$ to reveal the non-monotonicities in λ. In (b), vertical lines indicate the initial conditions used in (a). Also shown are the times taken under the deterministic dynamics (dashed lines) to get from the initial condition to within c/N of the stable fixed point; the order unity constant c is chosen to give a good description of the actual fixation times for initial conditions near the fixed point.

game by deterministic relaxation to the stable fixed point near $x = 1$, followed by noise-driven absorption. The increase of the fixation time with λ is shown in Fig. 5.4(b) and is qualitatively similar to the behaviour for the dominance game as plotted in Fig. 5.3(b).

For initial conditions with $x < x_0^\star$, the behaviour of the system and the resulting fixation time is more intricate, as shown in Fig. 5.5. Panel (a) demonstrates that the fixation time can now exhibit a *non-monotonic* dependence on the strength of memory loss λ, provided the starting point is sufficiently close to the location of the saddle-node bifurcation. The data in panel (b) show that the starting point has a non-trivial influence on fixation time.

This dependence on the initial condition x for $\lambda < \lambda_c$ can be understood as

follows. If x is smaller than the unstable (interior) fixed point at the given λ, deterministic relaxation will be to the stable fixed point at *lower* x, and activation from there will accordingly be to $x = 0$ rather than $x = 1$. A more detailed analysis for large N can be found in the appendices. This shows that close to the bifurcation, activation towards $x = 0$ is slower – exponentially in N – than across the barrier to the stable fixed point at large x, so the system follows the latter route and eventually reaches $x = 1$. We emphasize that this is a non-trivial prediction for the dynamics in finite populations; it cannot be deduced from the deterministic Sato-Crutchfield dynamics..

Moving beyond the bifurcation ($\lambda > \lambda_c$), the situation is simpler again. For sufficiently large N one predicts fixation by relaxation directly to the stable fixed point close to $x = 1$, and activation to $x = 1$ from there. In Fig. 5.6(a) one can see that the system relaxes to the stable fixed point close to $x = 1$ following the deterministic dynamics, then fixation occurs by activation. For small N and close to the bifurcation threshold, the system might initially stay in a region of relatively weak deterministic flow (see Fig. 5.6(b)). A detailed analysis of this phenomenon is deferred to the appendices.

5.3.5 Comparison with replicator-mutator dynamics

The effect of the entropic term in the Sato-Crutchfield equations is

akin to that of mutation in evolutionary processes. Such mutation dynamics is discussed in [57, 66], for example. Both mutation and entropic terms describe forces that act towards the centre of strategy space and drive the population away from states in which one species (or one idea) dominates, and we here include a brief comparison. We choose the replicator-mutator equation of the form discussed in [7]

$$\dot{x} = \left(1 - \frac{u}{2}\right) x(1-x)(\pi_1 - \pi_2) - \frac{u}{2}\left(x - \frac{1}{2}\right), \tag{5.24}$$

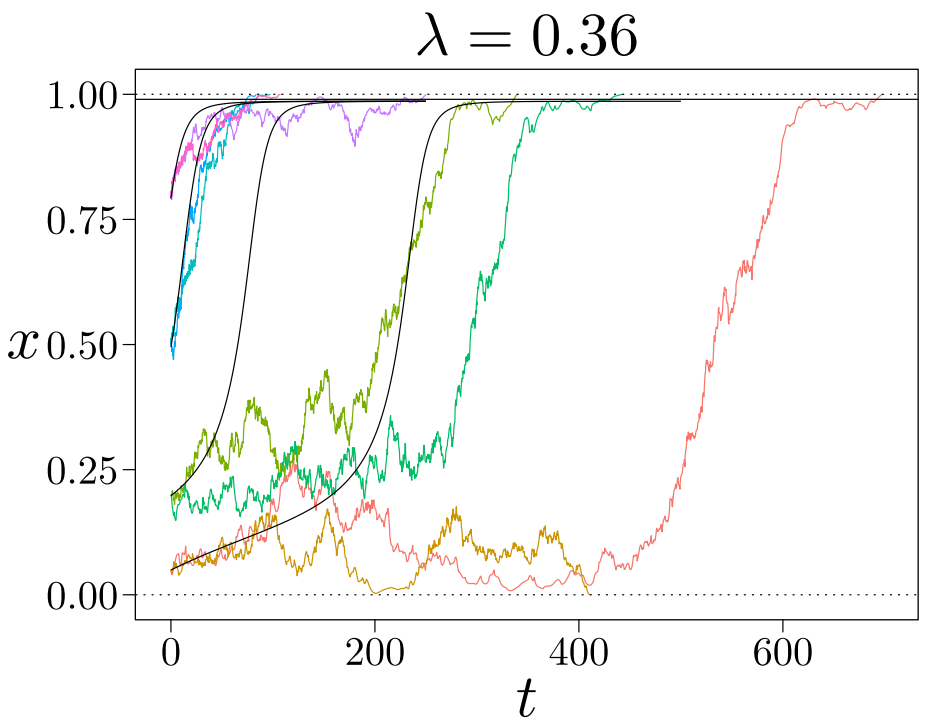

Figure 5.6: Sample trajectories of a coordination game, for different initial conditions; $\Gamma = 0.1$, $\lambda = 0.36$ and $N = 1000$ (coloured curves). The black curves show the trajectory for the deterministic dynamics (5.11) starting from the same set of initial conditions. The trajectories follow the deterministic dynamics fairly closely for initial conditions $x = 0.5$ and 0.8. For initial condition $x = 0.2$, fluctuations determine how fast the system escapes from the initial region of relatively weak deterministic flow. For initial $x = 0.05$, this effect is even stronger. One of the two trajectories shown also illustrates direct activation, to fixation at $x = 0$, against the deterministic flow.

where $u > 0$ is the mutation rate. In order to compare the effects of mutation with those of memory loss in the learning process, we show the bifurcation diagrams of the replicator-mutator dynamics along with those of Sato-Crutchfield learning in Fig. 5.7, for the three classes of symmetric games we have considered. The main difference between the two flows is that Sato-Crutchfield dynamics has additional fixed points at $x = 0$ and $x = 1$. As these are unstable for $\lambda > 0$, they do not lead to qualitative differences in the long-time

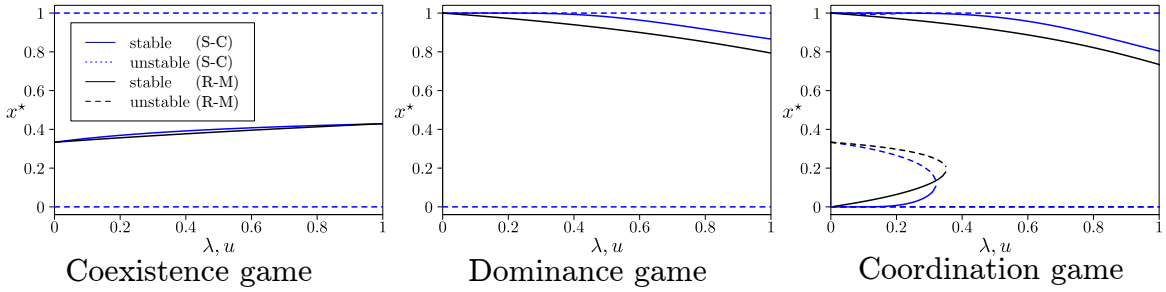

Figure 5.7: Fixed point diagrams of Sato-Crutchfield (S-C) learning (5.21), (blue and red lines), and replicator-mutator (R-M) dynamics (5.24) (black lines) for our three types of symmetric 2×2 games. The full black lines show the stable fixed points of the replicator-mutator dynamics and the dashed line its unstable fixed points.

deterministic dynamics. However, for finite N the difference is significant: replicator-mutator dynamics does not have absorbing states, so the question of fixation does not arise.

5.4 Asymmetric games and multiple populations of ideas

5.4.1 Birth-death dynamics for multiple populations of ideas

In this section we extend the stochastic dynamics for populations of ideas to games with multiple populations. We focus on the simplest case of two-player two-strategy games, though the approach easily extends to more general games. Our starting point are the Sato-Crutchfield equations (5.10), which simplify

to

$$\dot{x}^A = \Gamma x^A(\pi_1^A - \pi^A) - \Gamma\lambda x^A \left(\ln x^A + s^A\right), \tag{5.25a}$$

$$\dot{x}^B = \Gamma x^B(\pi_1^B - \pi^B) - \Gamma\lambda x^B \left(\ln x^B + s^B\right), \tag{5.25b}$$

where $\pi_1^A = a_{11}x^B + a_{12}(1-x^B)$, $\pi_2^A = a_{21}x^B + a_{22}(1-x^B)$, with analogous expressions for π_1^B and π_2^B. We have also written $\pi^A = x^A\pi_1^A + (1-x^A)\pi_2^A$, and $s^A = -\left[x^A \ln x^A + (1-x^A)\ln(1-x^A)\right]$. Similar definitions apply to π^B and s^B. The variable x^A denotes the probability with which player A chooses their action 1 and similarly for x^B.

The stochastic evolutionary dynamics now occurs in two finite populations of ideas, one for either player, each consisting of N individuals. We write n for the number of ideas of type 1 in population A, and similarly m for the number of ideas of type 1 in population B. The dynamics is defined by the rates for birth-death transitions in population A, $(n,m) \to (n \pm 1, m)$,

$$T^{A+}_{(n,m)} = \frac{1}{2}\frac{n(N-n)}{N}\left[1 + \Gamma\left(\pi_1^A - \pi_2^A - \lambda \ln\frac{n}{N-n}\right)\right], \tag{5.26a}$$

$$T^{A-}_{(n,m)} = \frac{1}{2}\frac{n(N-n)}{N}\left[1 + \Gamma\left(\pi_2^A - \pi_1^A - \lambda \ln\frac{N-n}{n}\right)\right], \tag{5.26b}$$

and analogous rates for transitions $(n,m) \to (n, m \pm 1)$ in population B

$$T^{B+}_{(n,m)} = \frac{1}{2}\frac{m(N-m)}{N}\left[1 + \Gamma\left(\pi_1^B - \pi_2^B - \lambda \ln\frac{m}{N-m}\right)\right], \tag{5.27a}$$

$$T^{B-}_{(n,m)} = \frac{1}{2}\frac{m(N-m)}{N}\left[1 + \Gamma\left(\pi_2^B - \pi_1^B - \lambda \ln\frac{N-m}{m}\right)\right]. \tag{5.27b}$$

The two-population birth-death dynamics has four absorbing states, $(n,m) = (0,0)$, $(0,N)$, $(N,0)$, (N,N) in finite populations. In the limit $N \to \infty$ and writing $x^A = n/N$ as well as $x^B = m/N$, this process leads to the deterministic two-population Sato-Crutchfield equations (5.25a) and (5.25b).

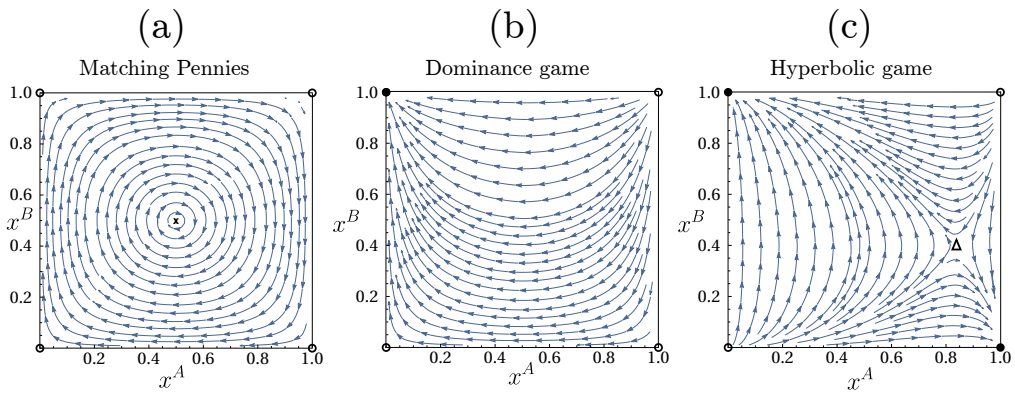

Figure 5.1: Two-population replicator flows for different types of 2×2 asymmetric games: (a) the Matching Pennies game is a zero-sum game; the replicator dynamics has a conserved quantity and exhibits cyclic trajectories. The game (b) has one pure-strategy fixed point while (c) has a hyperbolic fixed point. Stable fixed points are labeled by full dots, saddles (fixed points with one unstable and one stable direction) by triangles, unstable fixed points (two unstable directions) by empty dots and finally cyclic fixed points (whose Jacobian eigenvalues are purely imaginary) by a cross.

5.4.2 Examples of two-player two-strategy asymmetric games

We now study the corresponding fixation properties, focusing on a few key examples of asymmetric two-player games, chosen from the different categories of possible two-population replicator flows [50]: (i) the so-called Matching Pennies game, also known as Dawkin's Battle of the Sexes [34]; (ii) games in which one player has an action that strictly dominates the alternative action; and (iii) games in which the replicator flow has a hyperbolic interior fixed point. The three cases are illustrated in Fig. 5.1.

Matching Pennies game. This game is represented by the following payoff bi-matrix

$$\mathbf{A} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{5.28}$$

In addition to the trivial fixed points at the corners of phase space the replicator dynamics ($\lambda = 0$) has the fixed point $\mathbf{x}^\star = (x^A, x^B) = (0.5, 0.5)$. Trajectories that start elsewhere will form closed periodic orbits around the fixed point as shown in Fig. 5.2(a). Fixation in one of the four corners in *finite* populations will therefore be due to radial diffusion. Diffusion distances generally grow as $\sqrt{Dt}$. As the diffusion constant is $D \sim 1/N$ in our case, covering a radial distance of order unity to reach one of the two corners requires time $t \sim N$. This linear growth of fixation time with population size is shown in Fig. 5.2(d).

The effect can be seen as an analogue of the trapping in regions of low flow discussed in the anendices, but here the (radial) flow is zero over an extended region rather than at a single point, causing a stronger fixation time growth (N versus $\ln N$) with population size.

As soon as one has nonzero memory loss λ, the point $\mathbf{x}^\star$ becomes an attractor of the dynamics, with the whole state space as basin of attraction as shown in Fig. 5.2(b). As before, fixation will therefore proceed along the sequence of relaxation to this fixed point followed by activation to one of the absorbing states. The activation phase again requires a time scaling exponentially with the population size N. This change in scaling is clear by comparing Figs. 5.2(d) and (e) and emphasizes that the addition of the entropic term in the fitness has qualitative consequences for the fixation dynamics. The sample trajectories in Fig. 5.3 further illustrate this.

When λ becomes large, the flow and hence the activation barrier becomes proportional to λ to leading order, producing fixation times that scale exponentially with λ as can be seen in Fig. 5.2(c).

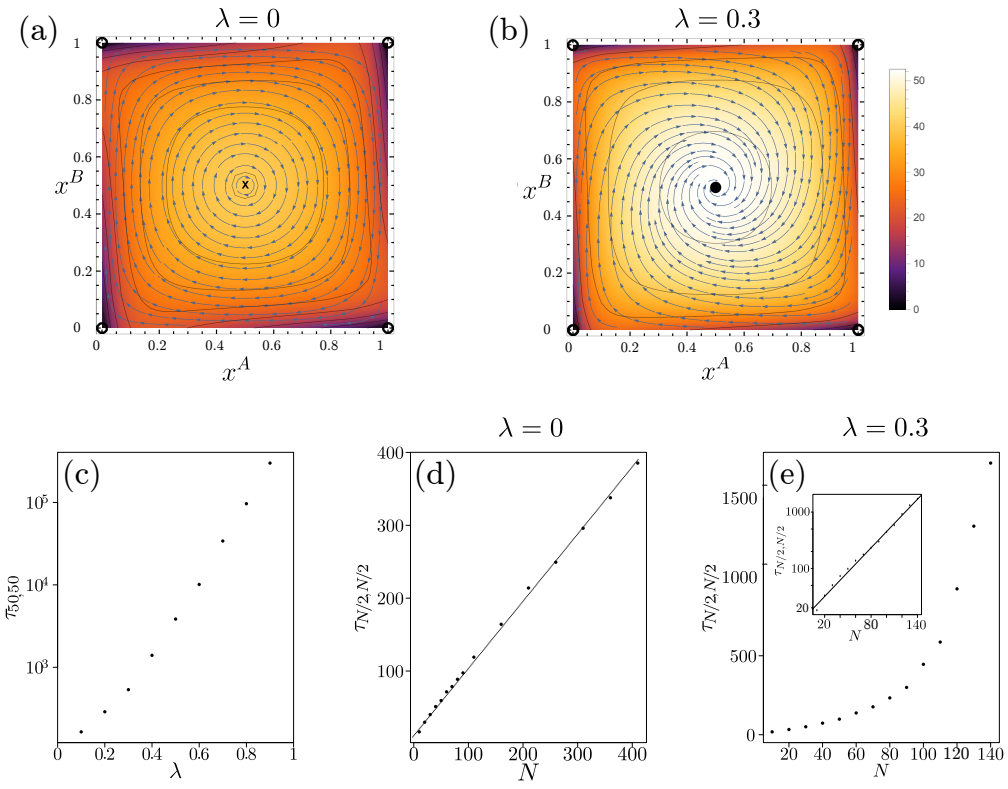

Figure 5.2: Matching Pennies game. (a,b) Flow under deterministic Sato-Crutchfield learning for $\lambda = 0$ and $\lambda = 0.3$, respectively. Overlaid is a heat map indicating the fixation time as a function of the starting point (obtained from the backward master equation [74] for a system of size $N = 30$); (c) Fixation time from simulations as a function of λ, for population size $N = 100$ and $(n, m) = (N/2, N/2)$ as initial condition. Panels (d, e) show fixation time $\tau_{N/2,N/2}$ against N for $\lambda = 0$ and $\lambda = 0.3$, respectively. Panel (d) shows linear scaling of fixation time with N (solid line) consistent with fixation by radial diffusion, whereas panel (e) displays approximately exponential scaling (see log-linear plot in inset) as fixation now requires activation against the flow.

Dominance game. An example of this case is defined by the payoff structure

$$\mathbf{A} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{5.29}$$

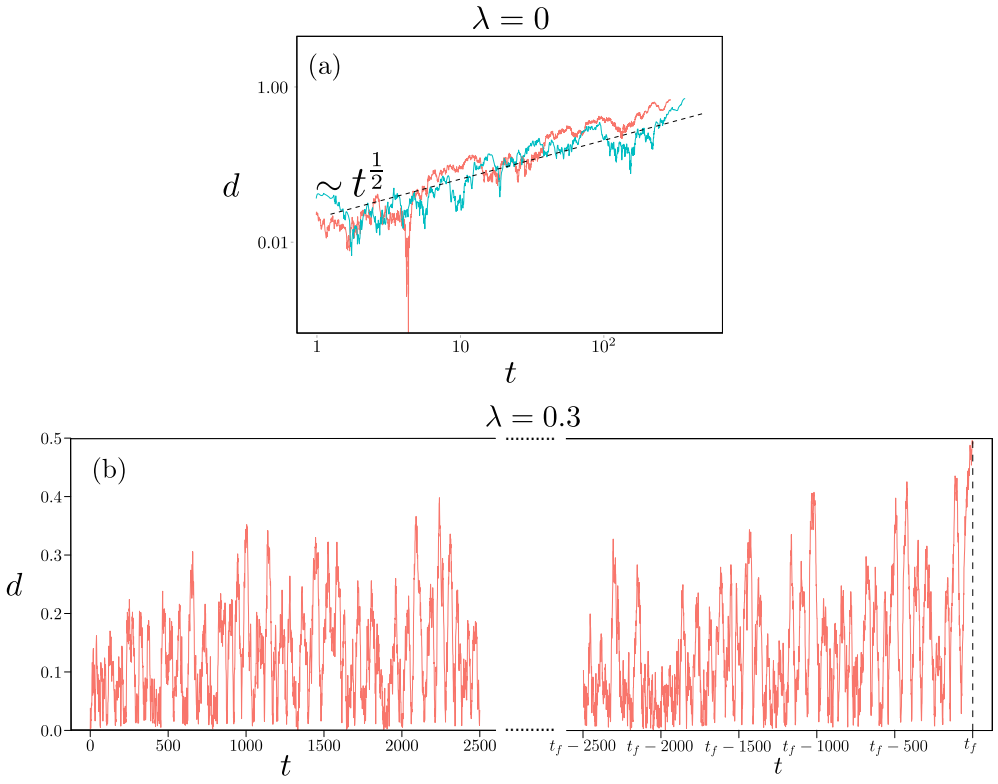

Figure 5.3: Matching Pennies game. Sample trajectories in a population of size $N = 500$ and with $\Gamma = 0.1$, for (a) $\lambda = 0$ and (b) $\lambda = 0.3$. We show the distance d of (x^A, x^B) from the fixed point at $(0.5, 0.5)$ versus time t in log-linear scale, to focus on the radial motion. Note the difference between diffusive dynamics in (a) – the dashed line shows the expected power law $1/2$ for a diffusive process – and activation in (b). For the latter we plot the beginning of the trajectory, showing how the system reaches a metastable steady state where it fluctuates around the centre of the state space ($d = 0$), and on the right the end of the fixation trajectory where a fluctuation takes the system to one of the four absorbing states at time t_f.

Its Sato-Crutchfield dynamics for $\lambda = 0$ has four fixed points in the corners of the state space, one of which is stable. Fixation will then typically proceed by deterministic relaxation to this fixed point. For infinite N this would take infinite time as the approach to the fixed point is exponential. At finite N, one expects that fixation takes place once this exponential approach gets within

distance $1/N$ – the grid spacing in the (x^A, x^B)-plane – of the fixed point. The fixation time should then scale logarithmically with N; the data in Fig. 5.4(d) are consistent with this.

As the memory-loss parameter λ is increased from zero, the stable fixed point moves continuously towards the centre of the state space, with all four corners then unstable fixed points. (There are also two additional saddle points on the boundary near the original stable fixed point.) Fixation will take place by relaxation followed by activation, resulting in exponential growth of fixation times with N (Fig. 5.4(e)) and, at large λ, also with λ (Fig. 5.4(c)). The sample trajectories in Fig. 5.5 illustrate the qualitative differences between the fixation dynamics for $\lambda = 0$ and $\lambda > 0$.

Hyperbolic game. An example of this class of games is given by the payoff matrices

$$\mathbf{A} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \tag{5.30}$$

For $\lambda = 0$, the Sato-Crutchfield dynamics has one saddle point in the interior of the state space, two stable fixed points in two opposite corners of the state space, and two unstable fixed points in the remaining corners; cf. Fig. 5.6. As for the dominance game, fixation will proceed by deterministic relaxation, leading to exponential approach to one of the two stable fixed points. Logarithmic growth with N of fixation times should again result, though we have not verified this explicitly.

Each of the two stable fixed points has its own basin of attraction. This is a new feature compared to the dominance game. For $N \to \infty$, the location in strategy space where fixation occurs will be entirely determined by which basin the system starts off in. For finite N, fluctuation effects will then make the choice of fixation location stochastic.

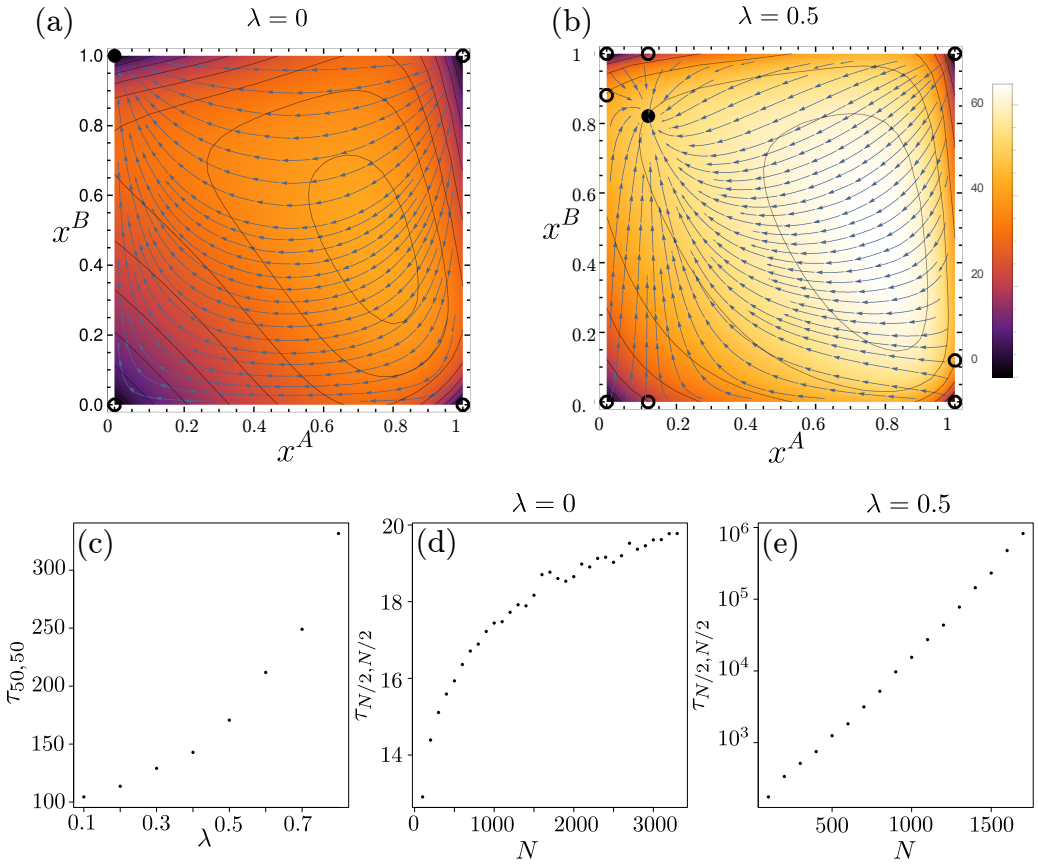

Figure 5.4: Dominance game. (a,b) Flow under deterministic Sato-Crutchfield learning for $\lambda = 0$ and $\lambda = 0.5$, respectively. Overlaid is a heat map indicating the mean fixation time as a function of the starting point (obtained from the backward master equation for a system of size $N = 30$); (c) Fixation time from Gillespie simulations as a function of λ, for population size $N = 100$ and $(n, m) = (N/2, N/2)$ as initial condition. (d, e) Fixation time $\tau_{N/2,N/2}$ against N for $\lambda = 0$ and 0.5, respectively. The fixation time in (d) exhibits logarithmic scaling with N resulting from the exponential approach to the stable fixed point. The scaling of the fixation time in (e) is approximately exponential with N because fixation involves activation.

With increasing λ, the two stable fixed points in the corners move to the interior of the state space. At a critical value λ_c, these two fixed points merge with the saddle point into a single stable fixed point. (This is the consequence of

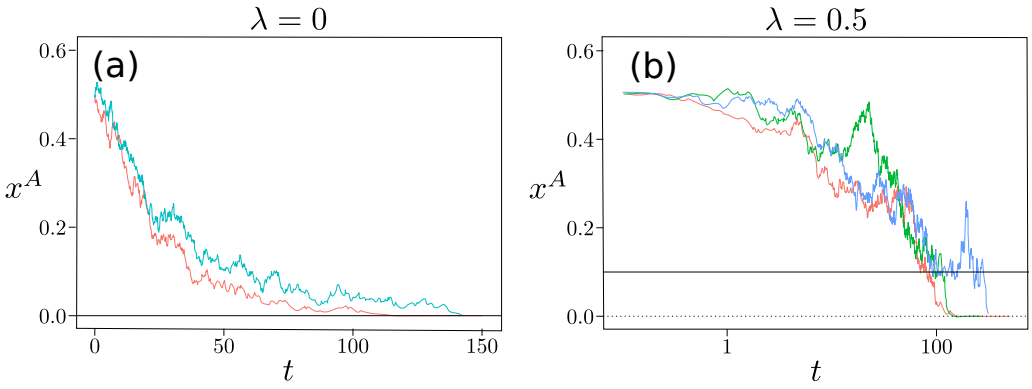

Figure 5.5: Dominance game. Sample trajectories in a population of size $N = 500$ and with $\Gamma = 0.1$, for (a) $\lambda = 0$ and (b) $\lambda = 0.5$. We show x^A against time (linear axis in (a), logarithmic axis in (b)). The full and dashed horizontal lines show the x^A-coordinate of the stable and unstable fixed points of the deterministic dynamics, see also Fig. 5.4.

a symmetry in our payoff matrices; without this, the saddle would annihilate with one stable fixed point and the other would survive.) The presence of this bifurcation would suggest, by analogy with the results for the coordination game, a non-monotonic dependence of the fixation time on λ near λ_c. Presumably the values of N required to see this will be large again, however, and we were unable to reach them in the two-population case with reasonable computational effort. Nonetheless, Fig. 5.7 illustrates clearly that as λ varies, the different fixed point structures of the deterministic dynamics cause qualitative changes in the fixation trajectories.

5.5 Summary and outlook

We have interpreted learning in games as a pairwise comparison process within a population of ideas. In the limit of large population size, the dynamics is described by the deterministic Sato-Crutchfield equations. While these equations for learning have been widely studied, there has (to our knowledge) not

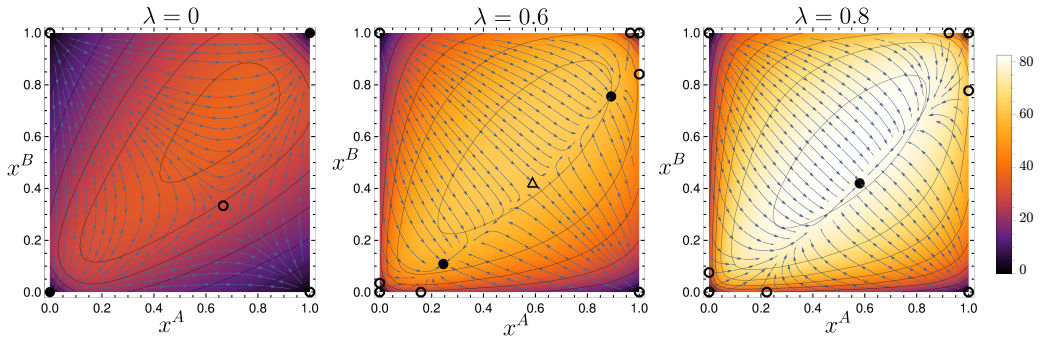

Figure 5.6: Hyperbolic game. Flow under deterministic Sato-Crutchfield learning for $\lambda = 0$, 0.6 and 0.8, respectively. Overlaid is in each panel a heat map showing the mean fixation time as a function of starting point in a system of size $N = 30$. The three chosen values of λ show different fixed point structures as indicated by the symbols.

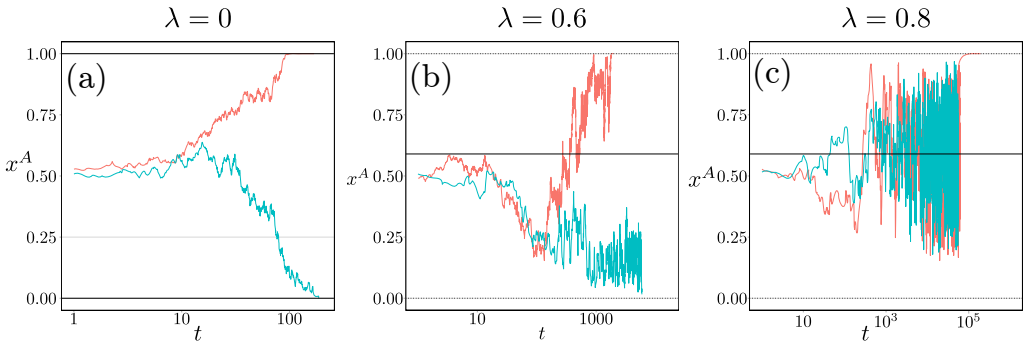

Figure 5.7: Hyperbolic game. Sample trajectories in a population of size $N = 500$ and with $\Gamma = 0.1$, for (a) $\lambda = 0$, (b) 0.6 and (c) 0.8, in the same representation as in Fig. 5.5. Panel (a) shows relaxation to the region around the saddle point, with fluctuations then determining at which boundary fixed point fixation occurs. The trajectories in (b) start similarly but then are driven to one of two *interior* stable fixed points, from which fixation proceeds by activation to the nearest boundary. In (c), all trajectories go to the single interior fixed point, from which fixation by activation occurs to one of two boundary fixed points (top right and bottom left in right-hand panel of Fig. 5.6).

been any systematic derivation from a birth-death process in finite populations. Such individual-based foundations are only available for simpler replicator (or

replicator-mutator) dynamics [99, 100, 7]. We fill this gap by defining such an individual-based process in a finite population of ideas. The construction in Sec. 5.3.3 and 5.4.1 involves augmenting the standard fitness function by a term proportional to the information content $(-\ln x_i)$ of species i. While the behaviour of deterministic Sato-Crutchfield learning in continuous time is fairly similar to the outcome of replicator-mutator dynamics in infinite populations, there are marked differences between their stochastic representations in finite systems. Mutation processes prevent fixation or extinction, but these phenomena can and will occur in finite populations of ideas, even at non-zero memory loss.

In order to develop some intuition for the general phenomena that can occur in finite populations of ideas we first studied three types of symmetric games (Sec. 5.3.4). We focused on the dependence of the fixation dynamics on the size of the population and on the memory-loss parameter λ. In our interpretation this latter parameter becomes the strength of the preference for rare ideas. The variety of different behaviours observed could be understood by decomposing the fixation dynamics into a sequence of elementary events, such as relaxation to stable fixed points and activation against the deterministic flow driven by demographic noise. We then broadened our analysis to include asymmetric two-player games (Sec. 5.4.2). Further features of the dynamics are then observed, such as fixation by diffusion when the relevant part of the dynamics is not opposed by the deterministic flow.

Most of our results are obtained from direct Gillespie simulations of the stochastic evolution of ideas, or from numerical solutions of the corresponding backward master equation. In the case of symmetric games we have complemented this with an analysis for large population size N (see Sec. 5.C of the appendices). This allows one to identify the dominant scaling of fixation times and reveals subtle effects that cannot be deduced from the fixed point structure of the dynamics (Sec. 5.B of the appendices). For asymmetric games there is in

general no mapping to noisy descent on an effective potential energy, because of the lack of detailed balance. However, as discussed e.g. by Bouchet et al. in [11], one should – in principle – be able to obtain fixation times for large N by using Freidlin-Wentzell large deviation theory. This is left to future work.

We think our work will enrich the mathematical theory of learning and evolutionary dynamics, providing a novel interpretation of learning in games with imperfect memory as a pairwise matching process between ideas. Our construction places the dynamics of learning in the context of stochastic population dynamics, and, we hope, it will encourage further studies of learning based on the established toolbox for evolutionary dynamics in finite populations.

Appendix

5.A Limits on birth-death description of Sato-Crutchfield learning

The parameters Γ and λ of the stochastic evolution of ideas we have defined need to be chosen so that all transition rates $T_n^{\pm}$ in Eq. (5.22) are non-negative. Except in the case of pure replicator dynamics ($\lambda = 0$), this gives constraints on the parameters that depend on population size N, though weakly. The reason is the logarithmic term in the fitness (5.19), which can get as large as $-\lambda \ln(1/N)$.

For fixed Γ the rates will only remain non-negative if $\lambda \leq \lambda_c$. One can compute a lower bound for λ_c. Firstly, all the transition rates will be positive if and only if the constraint

$$\left| \Delta\pi - \lambda \ln \frac{N-n}{n} \right| \leq \frac{1}{\Gamma} \tag{5.31}$$

is met for all $0 < n < N$. Since the quantity $\Delta\pi = \pi_1 - \pi_2$ varies linearly with n, it is bounded by $\Delta\pi(1)$ and $\Delta\pi(N-1)$. Applying the triangular inequality to (5.31) gives:

$$\left| \Delta\pi - \lambda \ln \frac{N-n}{n} \right| \leq \max\left(|\Delta\pi(1)|, |\Delta\pi(N-1)|\right) + \lambda \ln(N-1) \tag{5.32}$$

As a consequence, all transition rates are positive as long as $\max\left(|\Delta\pi(1)|, |\Delta\pi(N-1)|\right) + \lambda \ln(N-1) \leq 1/\Gamma$. This translates into

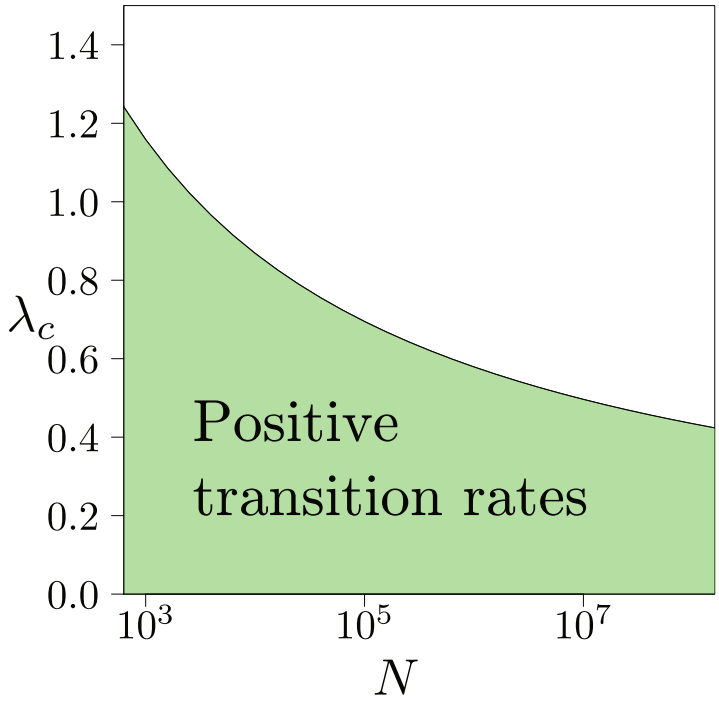

Figure 5.A.1: Lower bound on λ_c for a coexistence game as defined in section 5.3.4, for $\Gamma = 0.1$. The bound (black line) is inversely proportional to the logarithm of the population size N.

$$\lambda \leq \frac{1}{\ln(N-1)} \left(\frac{1}{\Gamma} - \max\left(|\Delta\pi(1)|, |\Delta\pi(N-1)| \right) \right) \tag{5.33}$$

The right-hand side therefore provides a lower bound on the critical value λ_c. This bound is plotted as a function of N in Fig. 5.A.1. While the bound goes to zero for $N \to \infty$, the inverse logarithmic dependence means it does so extremely slowly: the restriction on the allowed range of λ is therefore mild even for very large population sizes ($N \sim 10^8$ and beyond).

5.B Non-monotonicity of the fixation time in a coordination game

In Sec. 5.3.4, we studied fixation in a coordination game and observed that the fixation time is *non monotonic* in λ close to the bifurcation threshold λ_c, for *small* N. We will provide an explanation for this phenomenon by decomposing the dynamics leading to fixation into a sequence of elementary events. When N is small enough for activation times to be only moderate, beyond the bifurcation threshold, two additional effects come into play in addition to the relaxation and activation processes observed for large N: (i) direct activation: when starting near $x = 0$, a fluctuation (activation event) can drive the system straight to fixation at $x = 0$, even though the deterministic relaxation would take it in the other direction; (ii) trapping in regions near deterministic fixed points, where the net (deterministic) flow is low; deterministic relaxation times can then become comparable to activation times (precisely at such a fixed point, the deterministic relaxation time is in fact infinite as the flow vanishes). Finite populations will stay trapped in these regions of low deterministic flow for a long (but finite) time. This time will grow logarithmically with N as explained in this appendix, Sec. 5.D. Such regions exist at and near the bifurcation at λ_c, both for λ below and above λ_c.

The curve in Fig. 5.5(b) for $\lambda = 0.475$ shows the first effect: for small initial values of x, fixation times are rather low, as direct activation towards $x = 0$ is the dominant fixation mechanism. To the right of the maximum in the curve, on the other hand, we have fixation predominantly at $x = 1$. The fixation time here is, to a good approximation, given by the deterministic relaxation time to the stable fixed point close to $x = 1$, with the final activation to $x = 1$ being sufficiently fast to be sub-leading.

Accordingly, the sample trajectories in Fig. 5.6 show that the system moves

to the stable fixed point in a close-to-deterministic fashion, with fixation at $x = 1$ occurring shortly afterwards.

The second effect above contributes to the initial condition-dependence of the fixation time in Fig. 5.5. Here we are close enough to the bifurcation to have an extended region of low flow, causing a significant peak in the transition time curve. The low flow also makes fluctuation effects significant as explained above, and these cause deviations from the times predicted for purely deterministic relaxation. In Fig. 5.6, the sample trajectories that start from $n = 200$ ($x = 0.2$) illustrate this effect.

Finally, the low flow also makes direct activation to $x = 0$ fast, giving a larger region of initial x where this is the main fixation mechanism. As is clear from Fig. 5.5(b), the resulting movement of the peak in the fixation time is what causes the non-monotonic λ-dependence at fixed initial condition that is visible in Fig. 5.5(a). We refer to one of the two sample trajectories starting from $n = 50$ ($x = 0.05$) in Fig. 5.6 for an illustration of a direct activation event.

We note that the direct activation effects discussed above for the coordination game do occur also for coexistence and dominance games, with the same consequence that fixation times become small for initial conditions near $x = 0$. These other games do not have the additional features arising from the bifurcation in the coordination game, however, so do not show non-monotonic variation of the fixation time with λ.

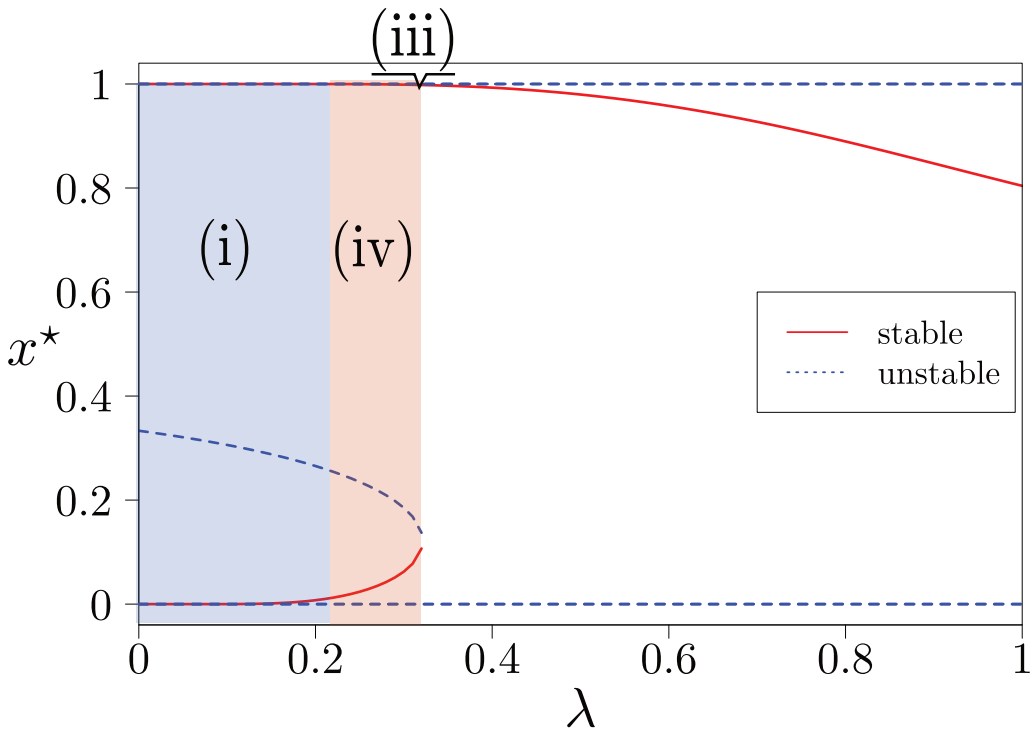

Figure 5.B.1: Different types of fixation dynamics in the coordination game with the payoff matrix of Fig. 5.1, superimposed onto the fixed point structure of Fig. 5.4(a). For values of λ below the bifurcation threshold, the potential formalism allows one to identify three different zones [(i), (iv) and (iii), with the latter covering only a very narrow λ-range] with qualitatively different fixation dynamics; see Fig. 5.C.2. Note that this subdivision into three zones cannot be deduced from the deterministic Sato-Crutchfield dynamics and its fixed point structure (Sec. 5.3.4) alone.

5.C Activation dynamics in stochastic evolution of ideas for symmetric games

Here, we explain how to obtain the large N-behaviour of activation times in our stochastic evolution for a population of ideas, and discuss the consequences for the fixation dynamics.

5.C.1 Kramers-Moyal expansion and effective potential

Our starting point is the dynamics defined by the transition rates (5.13a) and (5.13b). We have discussed in the main text how for $N \to \infty$ this leads to deterministic dynamics, here – by our construction – the Sato-Crutchfield equation (5.21). This can formally be derived from a Kramer-Moyal expansion to lowest order. In order to capture stochastic effects, one retains the first sub-leading order in the expansion. This is standard for evolutionary processes [100], and leads to an Itō stochastic differential equation of the form

$$\dot{x} = h(x) + \frac{1}{\sqrt{N}}\sigma(x)\xi(t), \tag{5.34}$$

where $\xi(t)$ is Gaussian white noise of unit variance, $\langle \xi(t)\xi(t')\rangle = \delta(t-t')$. For the birth-death process discussed in Sec. 5.3 one finds

$$h(x) = \Gamma x(1-x)\left(\pi_1(x) - \pi_2(x) - \lambda \ln\left(\frac{x}{1-x}\right)\right), \tag{5.35a}$$

$$\sigma(x) = \sqrt{x(1-x)}. \tag{5.35b}$$

Our aim is to use Eyring-Kramers theory [46], and so we map the above dynamics with multiplicative noise to one with additive noise. This is standard for systems with one degree of freedom, and is achieved by a change of variable from x to

$$y(x) \equiv \int_0^x \frac{\mathrm{d}x'}{\sigma(x')} = 2\arcsin(\sqrt{x}) \tag{5.36}$$

and conversely $x(y) = \sin^2(y/2)$. Translating the dynamics of x to one for y gives

$$\dot{y}(t) = \frac{h\left(x(y)\right)}{\sigma\left(x(y)\right)} - \underbrace{\frac{1}{2N}\frac{\sigma'(x(y))}{\sigma^2(x(y))}}_{\text{neglected}} + \frac{1}{\sqrt{N}}\xi(t) \tag{5.37}$$

The additional flow term with prefactor $\frac{1}{N}$ arises from the x-dependence of the original noise variance $\sigma^2(x)$. We will see shortly that this term can be neglected in determining the leading (exponential in N) scaling of activation times. The y-dynamics can now be written in the form

$$\dot{y}(t) = -\Gamma\frac{dV_y}{dy} + \frac{1}{\sqrt{N}}\xi(t) \tag{5.38}$$

with

$$V_y(y) = -\frac{1}{\Gamma}\int_0^y \mathrm{d}y' \; \frac{h(x(y'))}{\sigma(x(y'))} + \mathcal{O}(1/N) \tag{5.39}$$

Now that we have a standard Langevin equation with additive noise, Eyring-Kramers theory tells us that the time for an activated event, say from a stable fixed point y_1 to an unstable fixed point (barrier state) y_2 or to a boundary, scales as $\exp\{N\Gamma[V_y(y_2) - V_y(y_1)]\}$. It follows that the $\mathcal{O}(1/N)$ term in V_y will only contribute to the prefactor, which we are not considering here anyway; it can therefore be neglected. More importantly, if we translate back from y to x the potential takes the simple form

$$\begin{aligned} V(x) = V_y(y(x)) &= -\frac{1}{\Gamma}\int_0^x \frac{\mathrm{d}x'}{\sigma(x')}\frac{h(x')}{\sigma(x')} \\ &= -\int_0^x \mathrm{d}x' \left[\pi_1(x') - \pi_2(x') - \lambda \ln\left(\frac{x'}{1-x'}\right)\right] \end{aligned} \tag{5.40}$$

and activation times scale as

$$\tau \sim \exp\{N\Gamma[V(x_2) - V(x_1)]\} \tag{5.41}$$

This will be the basis for our further analysis. In particular, we will exploit that for large N, differences in activation barriers $V(x_2) - V(x_1)$ translate into exponentially different timescales, hence if there are competing processes the

one with the smaller activation barrier occurs first (with probability one as $N \to \infty$).

We add finally as a note of caution that the above Langevin analysis is valid for small Γ, where the rates for a transition $n \to n+1$ and its reverse are close to each other. Otherwise a more general approach is needed to determine activation timescales [45].

5.C.2 Generic symmetric two-strategy games

We can write down the potential $V(x)$ quite generically for a symmetric game where there are two actions to choose from. Inserting the explicit form of the payoffs (see Eq. (5.3a) and (5.3b)) into (5.40), one has, up to an umimportant additive constant,

$$V(x) = \tilde{v}\left(x - \frac{1}{2}\right) + \tilde{w}\left(x - \frac{1}{2}\right)^2 - \lambda s(x) \tag{5.42}$$

Here the entropy is $s(x) = -x\ln(x) - (1-x)\ln(1-x)$ as before, and we have introduced the abbreviations

$$\tilde{v} = \frac{a_{21} + a_{22} - a_{12} - a_{11}}{2} \tag{5.43a}$$

$$\tilde{w} = \frac{a_{12} + a_{21} - a_{11} - a_{22}}{2} \tag{5.43b}$$

For $\lambda = 0$ it is now easy to see the link to the three categories of symmetric game considered in Sec. 5.3.4, bearing in mind that all stationary points of $V(x)$ obey $h(x) = 0$, hence are fixed points of the dynamics. For $w > 0$ and $|v| < w$, $V(x)$ has a minimum in the relevant range $0 \le x \le 1$, and we have a *coexistence* game. For $w < 0$ and $|v| < |w|$, on the other hand, $V(x)$ has a maximum, corresponding to a *coordination* game. In the remaining cases, where $|v| > |w|$, $V(x)$ is monotonic for $x \in [0,1]$, so one has a *dominance* game.

To understand the effect of nonzero λ on $V(x)$, note that the function $-\lambda s(x)$ is convex. Hence for a coexistence game $V(x)$ continues to have a single minimum $x^\star$. A fixation trajectory will first relax to this minimum. The barrier to activation towards $x = 0$ is then $V(0) - V(x^\star)$, so fixation will occur there if this is lower than the corresponding barrier $V(1) - V(x^\star)$ for fixation at $x = 1$. In the opposite case, i.e. for $V(0) > V(1)$, fixation will occur at $x = 1$.

For a dominance game, the inclusion of the entropic term in $V(x)$ will create a single minimum $x^\star$ for any $\lambda > 0$, because the derivative $-\lambda s'(x)$ diverges to $\pm\infty$ at the two boundaries $x = 0$ and $x = 1$. The fixation dynamics then follows the same pattern as for a coexistence game.

5.C.3 Kramer-Moyal expansion for coordination games

The remaining case of coordination games is the most interesting, as the competition between the maximum in $V(x)$ at $\lambda = 0$ and the convex entropic term can create additional minima. We keep $\lambda > 0$ from now on and write

$$V(x) = \lambda \left[v\left(x - \frac{1}{2}\right) + w\left(x - \frac{1}{2}\right)^2 - s(x) \right] \tag{5.44}$$

with $v = \tilde{v}/\lambda$, $w = \tilde{w}/\lambda$. The shape of $V(x)$ is determined by these parameters, while λ only affects the overall scale of the activation barriers but not their relative size for different processes. We therefore drop the prefactor λ in the following.

For large v and w, corresponding to small λ at fixed $\tilde{v}$ and $\tilde{w}$, the entropic term is mostly negligible in $V(x)$. But its diverging derivative always dominates in $V'(x)$ when one is close enough to the boundaries, so must create two minima there. We denote their positions $x_1^\star$ and $x_2^\star$, respectively, and that of the

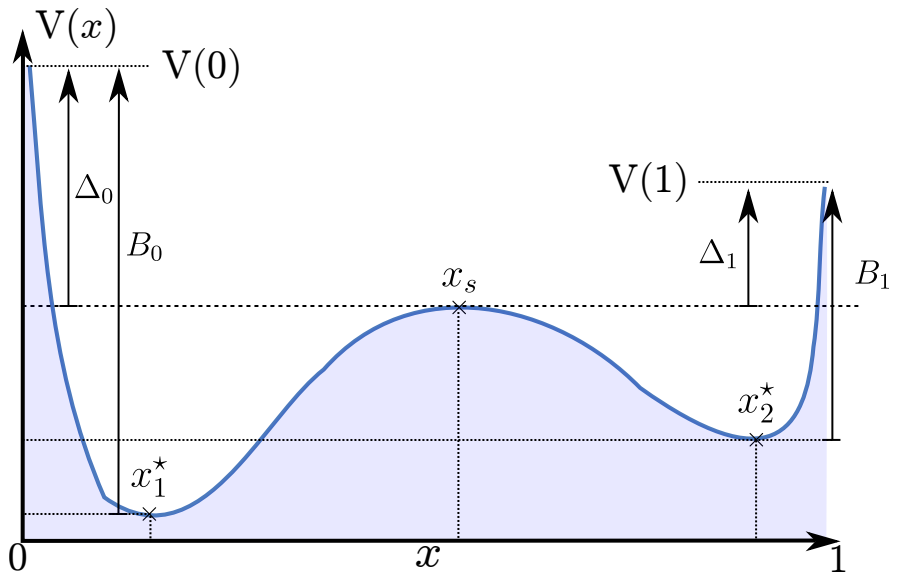

Figure 5.C.1: Graphical representation of the definitions of $\Delta_0 = V(0) - V(x_s)$, $\Delta_1 = V(1) - V(x_s)$, $B_0 = V(0) - V(x_1^\star)$ and $B_1 = V(1) - V(x_2^\star)$.

intermediate maximum by x_s. We also introduce

$$\begin{aligned} \Delta_0 &= V(0) - V(x_s), \qquad & \Delta_1 &= V(1) - V(x_s), \\ B_0 &= V(0) - V(x_1^\star), \qquad & B_1 &= V(1) - V(x_2^\star) \end{aligned} \tag{5.45}$$

as illustrated in Fig. 5.C.1. As v and w change, so will the values of these barrier parameters. In particular, the signs of Δ_0 and Δ_1 determine qualitatively the kind of fixation dynamics that the system will exhibit. The regime where Δ_0 and Δ_1 have *different* signs is subdivided further according to their relation to the barriers B_0 and B_1. A graphical summary is given in Fig. 5.C.2 and discussed further below. Fig. 5.C.3 shows the resulting phase diagram in the (v, w)-plane, and summarizes to what extent fixation probabilities and fixation times depend on initial conditions in each of the four regimes. Note that when w gets too close to zero, or $|v|/|w|$ becomes too large, a maximum and a minimum of $V(x)$ can merge in a bifurcation. In the single minimum regime beyond this, the fixation dynamics becomes simple again and has the same features as for coexistence and dominance games. The arrow in Fig. 5.C.3 shows how the various regions of the diagram are traversed when λ is increased at fixed $\tilde{v}$ and $\tilde{w}$, i.e. for fixed payoffs. In Fig. 5.B.1 we plot over what λ-ranges $V(x)$

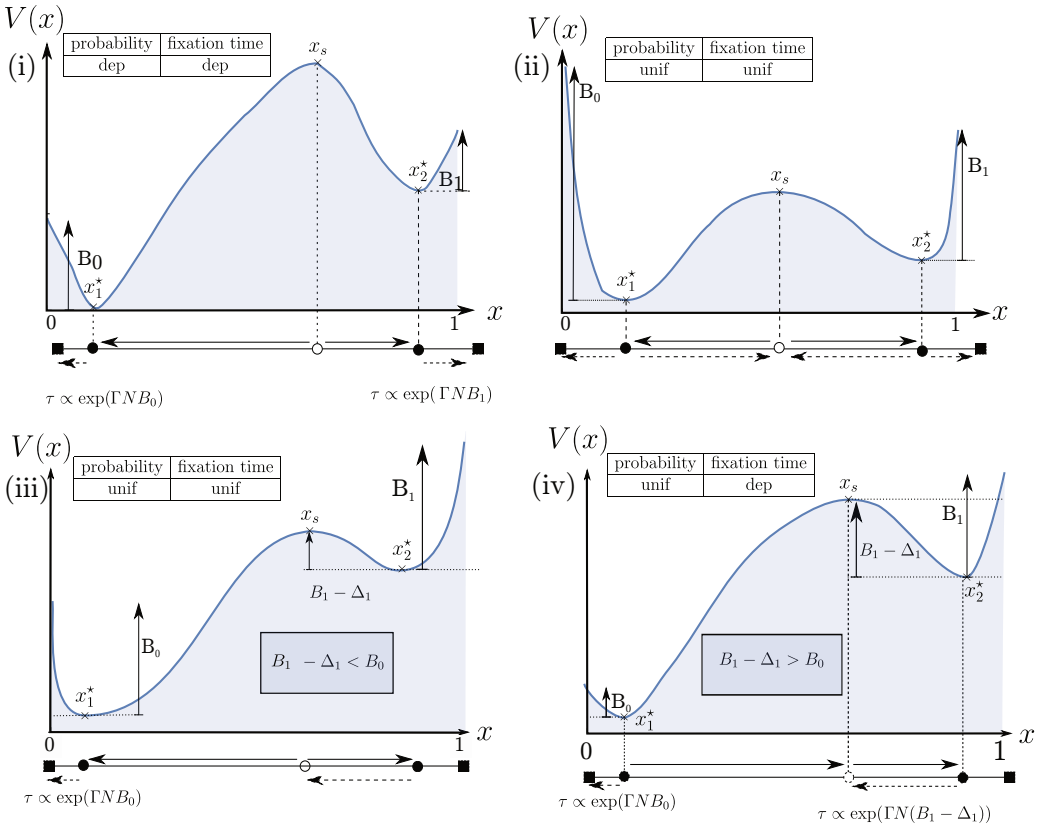

Figure 5.C.2: Schematic of the shape of the potential $V(x)$ in the four different classes of coordination games. Arrows on the bottom of each panel represent deterministic relaxation paths that occur during fixation (full lines) as well as activated events driven by fluctuations (dashed lines). The legends indicate whether the large N-fixation probability and fixation time depend on the initial position x, or are uniform in x.

has the shapes (i), (iii) and (iv), respectively, in the specific example game of Sec. 5.3.4. The λ-range for shape (iii) is too small to see in that figure, however.

Fig. 5.C.2(a) shows the simplest case $\Delta_0, \Delta_1 < 0$. Here depending on its initial condition, the system will first relax to one of the minima of the potential, say $x_1^\star$. Then because $\Delta_0 < 0$ the barrier for activation to $x = 0$ is smaller than for activation to the maximum x_s. For large N – which we always assume in the

following discussion – then with probability one the former process is the first to happen: fixation occurs at $x = 0$. Similarly if the initial relaxation goes to $x_2^\star$ because the system started at $x > x_s$, fixation will occur at $x = 1$. The fixation probability at 0 is therefore a step function of the initial condition x, dropping from one to zero at $x = x_s$. The fixation time changes similarly with initial condition, from $\exp[N\Gamma B_0]$ for $x < x_s$ to $\exp[N\Gamma B_1]$ for $x > x_s$.

The opposite case of $\Delta_0, \Delta_1 > 0$ is illustrated in Fig. 5.C.2(b). Here once the system has landed in either of the two minima, it will be able to reach the maximum separating these minima much faster than a boundary. As a result the system will make many "trips" between the two minima and effectively equilibrates across them, forgetting its initial condition. One can show that fixation will then eventually occur as if the system only had a single potential minimum at the *lower* of the two local potential minima, and will accordingly take place at the boundary with the lower value of V.

Finally there is the case where Δ_0 and Δ_1 have opposite signs, e.g. $\Delta_1 > 0$, $\Delta_0 < 0$ as shown in Fig. 5.C.2(c,d). If the system starts out of $x < x_s$, we have the same case as (a) above: deterministic relaxation to $x_1^\star$ followed by fixation at $x = 0$ on a timescale set by the barrier B_0. Otherwise, the system will initially relax to $x_2^\star$ and then traverse the maximum at x_s: $\Delta_1 > 0$ ensures that activation to the maximum is exponentially faster than fixation at $x = 1$. After arrival at $x_1^\star$ the earlier sequence of processes is followed. Because fixation in both cases takes place at $x = 0$, the fixation probability is independent of the initial condition.

Whether the fixation time has such a dependence, on the other hand, depends on timescales. As Fig. 5.C.2(c,d) shows, the timescale for activation from $x_2^\star$ to x_s is set by the barrier $B_1 - \Delta_1$, while the timescale for fixation at $x = 0$ from $x_1^\star$ is set by B_0. If the former is smaller than the latter, as in Fig. 5.C.2(c), then even when the system initially relaxes to $x_2^\star$, the timescale for the overall

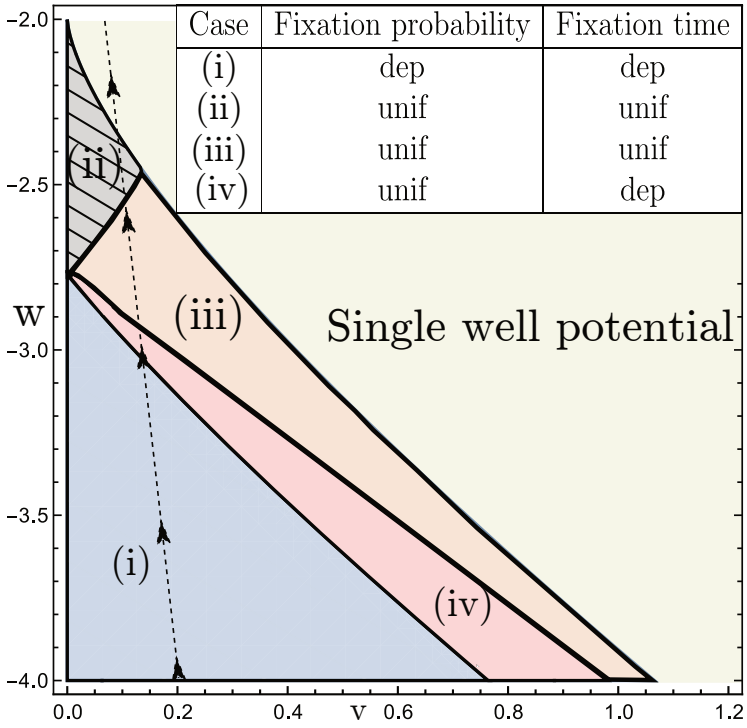

Figure 5.C.3: Phase diagram in the (v, w)-plane, indicating where the different shapes of $V(x)$ occur. These are explained in Fig. 5.C.2. The dotted arrow shows how the phase diagram is traversed at fixed $\tilde{v}$ and $\tilde{w}$ when λ is increased.

fixation trajectory will be given by B_0: it is therefore independent of the initial condition. In the converse case of Fig. 5.C.2(d), the system will take longer to reach fixation starting from $x > x_s$ because activation from $x_2^\star$ to x_s is much slower than fixation from $x_1^\star$. A typical fixation trajectory here will see the system spend almost all of its time near $x_2^\star$, before a fluctuation drives it across x_s to $x_1^\star$ and from there to $x = 0$.

5.D Fixation in regions of small flow

Here, we explain briefly why the noise-driven escape from the low-flow region around an unstable fixed fixed point takes a time scaling as $\ln(N)$.

Consider the linearized dynamics of a coordination (or other) game near an unstable fixed point. After the mapping to Langevin dynamics with additive noise, cf. (5.38), this can be written in the form

$$\dot{y} = \tilde{\mu}(y - y_0) + \frac{1}{\sqrt{N}}\xi(t) \tag{5.46}$$

with $\tilde{\mu} > 0$. Assuming that $y(0) = y_0$, a straightforward calculation then shows that the variance of $y(t)$ is:

$$\langle [y(t) - y_0]^2 \rangle = \frac{\exp(2\tilde{\mu}t) - 1}{2\tilde{\mu}N} \tag{5.47}$$

To have 'escape' from the unstable fixed point this needs to be of order unity; call this value c. Neglecting the -1 in the numerator then gives an escape time of order $t = \ln(2c\tilde{\mu}Nc + 1)/(2\tilde{\mu})$ which for large N becomes $\ln(N)/(2\tilde{\mu})$, establishing the promised logarithmic scaling with N. Note that while we have estimated the time for an escape to a distance of order unity in y-space, this is equivalent to an order unity distance in x-space as the mapping from x to y is smooth.

CHAPTER 6

Conclusion

6.1 Summary of the results

In this thesis we used large deviation methods to understand segregation and fixation in game theoretical models. Chaps. 2, 4 and 3 were concerned with the application of Freidlin-Wentzell theory to quantify the emergence of segregation for traders with an infinitely long memory. Chap. 5 dealt with the application of Kramers' rate theory, which is a simplified version of Freidlin-Wentzell theory for conservative dynamics, to study finite size quantities in birth-death processes.

The aim of Chap. 2 was twofold: (i) comparing the steady state of some EWA learning dynamics to a simple game theoretical model and (ii) studying how the emergence of segregation is influenced by the use of fictitious play. To fulfill aim (i) we built a simple aggregative game in Sec. 2.3, which turns out to have an infinity of Nash equilibria. We classified each of those Nash equilibria according to the type of strategies played by the traders (pure or mixed). In the rest of the chapter, we then linked the Nash equilibria to the actual long run outcome of EWA learning dynamics. We first investigated

the outcome of EWA learning dynamics when the fictitious play coefficient α is zero, traders have infinite memory $r \to 0$ and they best-respond to their preferences ($\beta \to \infty$). We found in this case that the steady state of EWA learning is a homogeneous mixed Nash equilibrium *i.e.* unimodal distribution of scores where all the traders play a mixed strategy. We then extended our analysis and studied the long run outcome of the EWA learning dynamics when the limits $\alpha \to 0$ and $\beta \to \infty$ are taken simultaneously. As shown in Fig. 2.5, depending on how the limit is taken, the corresponding long run outcome of the EWA learning dynamics will have qualitatively different properties (see Fig. 2.3). Those include homogeneous-mixed states where all traders within a classe randomize in the same way as well as heterogeneous-pure states where traders split into two groups each choosing a market deterministically. Along with these standard types of Nash equilibria, we also found a heterogeneous mixed steady states where traders splits into groups but not all groups play deterministically.

In Chap. 3 we used the equal action formalism developed in Chap. 2 to study strong segregation of traders with fixed buy/sell preferences across more than two markets. Motivated by the wide variety of qualitatively different scores distributions that we observed in multi-agent simulation (see Fig. 3.1), in Secs. 3.3 and 3.4 we explored segregation across three markets for different ranges of parameters. Sec. 3.3 was dedicated to the analysis of segregation across three fair markets, *i.e.* with $\theta_1 = \theta_2 = \theta_3 = 0.5$; this is the only set of market biases for which we found segregation of both classes of traders into three groups. Sec. 3.4 was concerned with the exploration of different ranges of market parameters to understand the causes that drive segregation. Based on this study, we inferred two plausible triggers for the emergence of segregation across three markets: (i) *the closeness between the markets*, (ii) *the average volume of trade and the profit earned in this market.*

Chap. 4 was concerned with the existence of segregation in a population where

the intensity of choice, as well as the memory, are heterogeneous. In Sec. 4.2 we detailed our approach to model heterogeneity. The largest section of the chapter, Sec. 4.3, was concerned with the consequences of various sources of heterogeneity. In Sec. 4.3.1 we observed that adding the same fraction of traders with short memory in each class decreases the segregation thresholds, *i.e.* suppresses segregation, for the traders with a long memory. Specifically, we saw that the variations with respect to the fraction of fast traders are significantly larger for the strong than for the weak segregation threshold. Sec. 4.3.2 investigated the effects of adding heterogeneity in the intensity of choice and we observed that decreasing the inverse intensity of choice of fast traders decreases the segregation threshold of slow traders. Finally, Sec. 4.3.3 explored how the segregation thresholds vary when the fraction of traders with short memory is different within the two classes. Although the results showed some non-trivial structure, our results were in line with the general trend that adding traders with short memory lowers the segregation threshold (see Fig. 4.1). We believe that this behaviour arises because adding fast traders reduces the range of possible market conditions that slow traders can reach by self-organising (see Fig. 4.2).

In Chap. 5 we interpreted learning as a pairwise comparison process within populations of ideas. While for large populations, the dynamics is described by the Sato-Crutchfield equation, there has been (to our knowledge) no systematic derivation of the Sato-Crutchfield equation as the large size limit of a birth-death process. Our work aimed at filling this gap by defining a birth-death process that leads to the Sato-Crutchfield equation in the limit of a population of large size. To do so we constructed in Sec. 5.3.3 and 5.4.1 a birth-death process where the fitness of individuals was augmented by a term proportional to the information content of species i ($-\ln x_i$). One of the outcomes of this work was to show that although the replicator-mutator dynamics and the Sato-Crutchfield dynamics have fairly similar behaviour, their stochastic extensions have qualitatively different properties. We also investigated the finite size

quantities of the dynamics of the population of ideas in a number of symmetric and asymmetric games as a function of the memory-loss parameter λ.

6.2 Future work

6.2.1 Technical improvements

The robustness of the action minimization algorithm: The algorithm we used to study strong segregation phenomena in Chaps. 2 and 3 is suitable for our parameter regime. However, for small values of the fictitious play coefficient, such as the one we were interested in in Chap. 2, the algorithm can fail to find the minimal action path (*i.e.* it does not converge). This prevented us from investigating very small values of the fictitious play coefficient α. Therefore, a natural direction for future work would be to improve the robustness of the action minimisation algorithm for small α. To do so we see two possible directions, either (i) improving the algorithm used to minimise the discretised minimal action path or (ii) directly solving the Hamilton-Jacobi equation that has as its solution the minimal action path. The latter turns out to be challenging because it requires the solution of a boundary value problem that is generally tackled with the "shooting method" [78]. However, in preliminary numerical experiments we found that the shooting method was not robust enough, which lead us to the use of the discretized path instead. An alternative technique to perform the shooting method could be based on linearising in small changes to the initial condition for the optimal path, from which iterative corrections to these initial conditions can then be estimated. However, this technique is still being developed and the method was not yet robust enough to be used in the numerics for this thesis.

Assessing the segregation via the distribution of scores: The advantage of the action minimisation methods we deployed is that they enable one to assess the existence, or otherwise, of segregation in the large memory limit. However—as explained above—in certain parameter regimes, the action minimisation algorithm may not converge. In such situations, one can assess the existence of segregation by looking at the distribution of scores in multi-agent simulations. One possible measure employed by Alorić [2, 1] is the Binder cumulant [6]. This coefficient gives helpful information about the bimodality of the score distribution when the peaks are of equal size but are of less useful when the score distribution is composed of peaks with different size. Alternatively, the existence of segregation can be tested using the DIP test for unimodality [47]. DIP compares the distribution of scores with the best-fitted uni-modal curve and returns a number between zero and one describing the similarity of the two curves. However, we see two challenges in using the DIP test for unimodality: (i) during our early investigation this test did not manage to separate strong and weak segregation (ii) as opposed to the Binder cumulant, we do not have any intuition on how the results of the DIP test generalise in the large memory limit $r \to 0$. Thus, it would be interesting to develop a generalised version of the DIP test that is adapted specifically to distributions with narrow peaks.

Asymptotics for the fixation time in populations of ideas In Chap. 5 we calculate the fixation times for a large population of ideas undergoing a birth-death process. We observed that the fixation time of such populations grows exponentially with its size. It would be interesting to calculate this exponential growth rate with the memory size for both symmetric and asymmetric games. We have considered such an extension of our work and note the following:

- For symmetric games, one should use Kramers' rate theory. A challenge arises when applying this theory, however. One would need to go beyond

the Kramers-Moyal expansion. If not, the resulting calculations would lead to an overestimate of the exponential growth rate of the fixation time [45, 26].

- Since the dynamics of asymmetric games is not conservative, here one would need to use a version of Freidlin-Wentzell theory that generalizes Kramers' rate theory to non-conservative systems [38, 11].

6.2.2 Extensions of the model

In Chaps. 2, 3 and 4, we examined the existence or otherwise of segregation in different variations of a minimal model of double auction markets. A logical continuation of this work is to explore further variations of the model among which we would suggest the following.

Adding heterogeneity in the traders' preferences Building upon the work of Chap. 4, we could add additional heterogeneity in the behaviour of the traders. For example, we could consider more than two different memory lengths within the population of traders. In such a case, we expect the analysis of both the weak and strong segregation thresholds to be tractable as long as the orders of magnitude of the different timescales of the dynamics of the traders are different.

As a consequence of the large number of memory timescales, one might expect the payoff of traders to be non-monotonic with their memory as reported by Toth *et al.* in [98]. This monotonicity is related to the fact that highly informed traders can take advantage of the less informed ones but cannot take advantage of the uninformed traders who trade almost randomly.

Another exciting direction would be to have a population of traders that have

different exploration/exploitation tradeoffs. A straightforward way to implement such a behaviour would be to have traders with various intensities of choice β, building upon the model developed in Chap. 4.

Varying the learning mechanisms The automated traders set up by the organizers of the CAT game that inspired our work learned their preferences using numerous different learning algorithms [16]. Starting from this fact, we believe it would be interesting to vary the learning mechanisms of the traders. This learning dynamics can be subdivided into how the traders make two separate choices: (i) the selection of their preferred market and (ii) the value of the order an agent sends to the market.

In our current model, we chose to have *zero intelligence traders*, who do not learn the price of the orders they send to the market. This choice was motivated by the work of Gode and Sunders [44] who regard the "Market as a Partial Substitute for Individual Rationality". An alternative proposed by Cliff *et al.* is the Zero-Intelligence-Plus [27] learning algorithm (instead of zero intelligence [44]) as a potentially more accurate model traders' behaviour. Other possible rules for traders to set their order price can be found in [36] and [43].

One could also study the robustness of the model against variation of the market selection strategy (reinforcement learning) our traders use. A simplified approach would be to consider market selection as a multi-armed-bandit problem (MAB). Among the existing learning algorithms used to solve such problems one can cite for example ϵ-greedy strategy, and regret minimization. This list is not exhaustive and there exist many other market selection strategies. We refer the interested reader to a comprehensive review of MAB problems for more information [94, 22].

Another exciting direction for our study would be to implement trader-trader interactions in our model. In the vein of [31], we could, for example, consider

random networks of traders interacting with each other and seeing how the connectivity of the network influences the emergence of segregation.

Varying markets mechanisms In Chaps. 2, 3 and 4, we considered a simple model of discrete double auction markets. This made it possible to obtain analytical solutions for the segregation threshold in the large memory limit. However, it would be interesting to generalize our results to a wider range of markets. We could, for example, add transaction fees as in [91] and make the markets adaptive as in [81] or [16]. There are many other market mechanisms one could explore and we suggest [71, 77] as an overview of the possible alternative market mechanisms that one could implement.

In real double auction markets, the traders pay for the service provided by the market. This payment can come from the bid-ask spread, from additional fees charged for each transaction, etc. Those features are not yet implemented in our model, and they are good candidates for further extensions of our work as they could contribute to making the model more realistic. One could, for example, imagine that the market charges a fixed fee for each transaction, or a fee proportional to the profit traders make, etc.

Segregation in other aggregative games In Chap. 2, we observed segregation in a double auction market modelled as an aggregative game. We believe it would be exciting to see whether such results can be generalized to other types of aggregative games such as the minority game [28, 24] or the Cournot model for an oligopoly [33]. One advantage of those games is that their simplicity should make it possible to obtain analytical results for the emergence of segregation using methods similar to those used in Chaps. 2, 3 and 4.

Bibliography

[1] Aleksandra Alorić. *Spontaneous Segregation of Adaptive Agents in Auctions*. PhD thesis, King's College London, 2017.

[2] Aleksandra Alorić, Peter Sollich, Peter McBurney, and Tobias Galla. Emergence of cooperative long-term market loyalty in double auction markets. *PLOS ONE*, 11(4):1–26, Apr. 2016.

[3] Philipp Altrock and Arne Traulsen. Fixation times in evolutionary games under weak selection. *New Journal of Physics*, 11(1):013012, Jan. 2009.

[4] Tibor Antal and István Scheuring. Fixation of strategies for an evolutionary game in finite populations. *Bulletin of Mathematical Biology*, 68(8):1923–1944, 2006.

[5] W. Brian Arthur. Inductive reasoning and bounded rationality. *The American Economic Review*, 84(2):406–411, 1994.

[6] Kurt Binder. Finite size scaling analysis of ising model block distribution functions. *Zeitschrift für Physik B Condensed Matter*, 43(2):119–140, Jun. 1981.

[7] Alex Bladon, Tobias Galla, and Alan McKane. Evolutionary dynamics, intrinsic noise, and cycles of cooperation. *Physical Review E*, 81(6):066122, 2010.

[8] Daan Bloembergen, Daniel Hennes, Peter McBurney, and Karl Tuyls. Trading in markets with noisy information: an evolutionary analysis. *Connection Science*, 27(3):253–268, 2015.

[9] Richard Blythe and Alan McKane. Stochastic models of evolution in genetics, ecology and linguistics. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(07):P07018, 2007.

[10] Tilman Börgers and Rajiv Sarin. Learning through reinforcement and replicator dynamics. *Journal of Economic Theory*, 77(1):1–14, Nov. 1997.

[11] Freddy Bouchet and Julien Reygner. Generalisation of the Eyring–Kramers transition rate formula to irreversible diffusion processes. *Annales de l'Institut Henri Poincaré*, 17(12):3499–3532, 2016.

[12] Serena Bradde and Giulio Biroli. The generalized Arrhenius law in out of equilibrium systems. *arXiv preprint arXiv:1204.6027*, 2012.

[13] William Brock and Cars Hommes. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic Dynamics and Control*, 22(8-9):1235–1274, 1998.

[14] Guy Bunin, Yariv Kafri, and Daniel Podolsky. Large deviations in boundary-driven systems: Numerical evaluation and effective large-scale behavior. *Europhysics Letters*, 99(2):20002, 2012.

[15] Luis Cabral. Asymmetric equilibria in symmetric games with many players. *Economics Letters*, 27(3):205 – 208, 1988.

[16] Kai Cai, Enrico Gerding, Peter McBurney, Jinzhong Niu, Simon Parsons, and Steve Phelps. Overview of CAT: A market design competition. Technical report, Department of Computer Science, University of Liverpool, 2009.

[17] Kai Cai, Jinzhong Niu, and Simon Parsons. On the effects of competition between agent-based double auction markets. *Electronic Commerce Research and Applications*, 13(4):229–242, Jul. 2014.

[18] Colin Camerer. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press, Princeton, NJ, 2003.

[19] Colin Camerer and Teck Hua Ho. Experience-weighted attraction learning in normal form games. *Econometrica*, 67(4):827–874, 1999.

[20] Pierre Cardaliaguet. Notes on mean field games. Technical report, Université Paris-Dauphine, 2010. https://www.ceremade.dauphine.fr/~cardaliaguet/MFG20130420.pdf.

[21] Guilherme Carmona. Nash equilibria of games with a continuum of players. *SSRN Electronic Journal*, 2004.

[22] Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge University Press, Cambridge, UK, 2006.

[23] Anirban Chakraborti, Ioane Muni Toke, Marco Patriarca, and Frédéric Abergel. Econophysics review: II. agent-based models. *Quantitative Finance*, 11(7):1013–1041, 2011.

[24] Damien Challet, Matteo Marsili, and Yi-Cheng Zhang. *Minority Games: Interacting agents in financial markets*. Oxford University Press, Oxford, UK, 2004.

[25] Carl Chiarella and Giulia Iori. A simulation analysis of the microstructure of double auction markets. *Quantitative Finance*, 2(5):346–353, 2002.

[26] Claudia Cianci, Duccio Fanelli, and Alan McKane. WKB versus general-

ized van kampen system-size expansion: the stochastic logistic equation. *arXiv preprint arXiv:1508.00490*, 2015.

[27] Dave Cliff and Janet Bruten. Zero not enough: On the lower limit of agent intelligence for continuous double auction markets. *HP Laboratories Technical Report HPL*, 28(3):633, 1997.

[28] A. C. C. Coolen. *The mathematical theory of minority games*. Oxford University Press, Oxford, UK, 2005.

[29] Luis Corchón. Comparative statics for aggregative games the strong concavity case. *Mathematical Social Sciences*, 28(3):151 – 165, 1994.

[30] Augustin Cournot. *Recherches sur les principes mathématiques de la théorie des richesses*. L. Hachette, 1838.

[31] Luca Dall'Asta, Matteo Marsili, and Paolo Pin. Collaboration in social networks. *Proceedings of the National Academy of Sciences*, 109(12):4395–4400, 2012.

[32] Constantinos Daskalakis, Paul Goldberg, and Christos Papadimitriou. The complexity of computing a Nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259, 2009.

[33] Andrew Daughety. *Cournot oligopoly: characterization and applications*. Cambridge University Press, Cambridge, UK, 2005.

[34] Richard Dawkins. *The Selfish Gene*. Oxford University Press, Oxford, UK, 1976.

[35] Glenn Ellison and Drew Fudenberg. Knife-edge or plateau: When do market models tip? *Quarterly Journal of Economics*, 118(4):1249, 2003.

[36] Ido Erev and Alvin Roth. Predicting how people play games: Rein-

forcement learning in experimental games with unique, mixed strategy equilibria. *The American Economic Review*, 88(4):848–881, 1998.

[37] Henry Eyring. The activated complex in chemical reactions. *Journal of Chemical Physics*, 3(2):107–115, 1935.

[38] Mark Freidlin and Alexander Wentzell. *Random Perturbations of Dynamical Systems*. Springer, Berlin, 1998.

[39] Eric Friedman and Scott Shenker. Learning and implementation on the internet. Departmental working papers, Rutgers University, Department of Economics, 1998.

[40] Drew Fudenberg and David Levine. *The theory of learning in games*. MIT Press, Cambridge, MA, 1998.

[41] Tobias Galla. Intrinsic noise in game dynamical learning. *Physical Review Letters*, 103(19):198702, 2009.

[42] Tobias Galla and Doyne Farmer. Complex dynamics in learning complicated games. *Proceedings of the National Academy of Sciences*, 110(4):1232–1236, 2013.

[43] Steven Gjerstad and John Dickhaut. Price formation in double auctions. *Games and Economic Behavior*, 22(1):1–29, 1998.

[44] Dhananjay Gode and Shyam Sunder. Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of Political Economy*, 101(1):119–137, 1993.

[45] Peter Hänggi, Hermann Grabert, Peter Talkner, and Harry Thomas. Bistable systems: master equation versus Fokker–Planck modeling. *Physical Review A*, 29(1):371, 1984.

[46] Peter Hänggi, Peter Talkner, and Michal Borkovec. Reaction-rate theory: fifty years after Kramers. *Reviews of Modern Physics*, 62(2):251, 1990.

[47] John Hartigan and Pamela Hartigan. The DIP test of unimodality. *The Annals of Statistics*, 13(1):70–84, Mar. 1985.

[48] Matthias Heymann and Eric Vanden-Eijnden. Pathways of maximum likelihood for rare events in nonequilibrium systems: Application to nucleation in the presence of shear. *Physical Review Letters*, 100:140601, 2008.

[49] Teck-Hua Ho, Colin Camerer, and Juin-Kuan Chong. Self-tuning experience weighted attraction learning in games. *Journal of Economic Theory*, 133(1):177–198, 2007.

[50] Josef Hofbauer and Karl Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge, UK, 1998.

[51] Zi-Gang Huang, Ji-Qiang Zhang, Jia-Qi Dong, Liang Huang, and Ying-Cheng Lai. Emergence of grouping in multi-resource minority game dynamics. *Scientific Reports*, 2(1), Oct. 2012.

[52] Neil Johnson, Pak Ming Hui, Zheng Dafang, and Michael Hart. Enhanced winnings in a mixed-ability population playing a minority game. *Journal of Physics A: Mathematical and General*, 32(38):L427, 1999.

[53] Neil Johnson, Pak Ming Hui, Rob Jonson, and Ting Shek Lo. Self-organized segregation within an evolving population. *Physical Review Letters*, 82:3360–3363, Apr. 1999.

[54] Daniel Kahneman. Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 93(5):1449–1475, Nov. 2003.

[55] Ian Kash, Eric Friedman, and Joseph Halpern. Multiagent learning in large anonymous games. *Journal of Artificial Intelligence Research*, 40:571–598, 2011.

[56] Michael Keisers. *Learning against Learning: Evolutionary Dynamics of Reinforcement Learning Algorithms in Strategic Interactions*. PhD thesis, Maastricht University, 2012.

[57] Nathalia Komarova. Replicator–mutator equation, universality property and population dynamics of learning. *Journal of Theoretical Biology*, 230(2):227–239, 2004.

[58] Hendrik Anthony Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940.

[59] Jean-Michel Lasry and Pierre-Louis Lions. Mean field games. *Japanese Journal of Mathematics*, 2(1):229–260, 2007.

[60] Ananth Madhavan. Consolidation, fragmentation, and the disclosure of trading information. *Review of Financial Studies*, 8(3):579, 1995.

[61] John Maynard-Smith and George Price. The logic of animal conflict. *Nature*, 246(5427):15–18, Nov. 1973.

[62] Roger McCain. Game theory and public policy, 2009.

[63] Marc Mezard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin Glass Theory and Beyond*. World Scientific Publishing Company, 1987.

[64] Paul Robert Milgrom. *Putting Auction Theory to Work*. Cambridge University Press, 2000.

[65] Tim Miller and Jinzhong Niu. An assessment of strategies for choosing

between competitive marketplaces. *Electronic Commerce Research and Applications*, 11:14–23, 2012.

[66] Mauro Mobilia. Oscillatory dynamics in rock–paper–scissors games with mutations. *Journal of Theoretical Biology*, 264(1):1–10, 2010.

[67] Mauro Mobilia and Michael Assaf. Fixation in evolutionary games under non-vanishing selection. *Europhysics Letters*, 91(1):10002, 2010.

[68] John Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49, 1950.

[69] John Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, 3rd. edition, 1953.

[70] Abraham Neyman. Bounded complexity justifies cooperation in the finitely repeated prisonner dilemma. *Economics Letters*, 19:227–229, 1985.

[71] Jinzhong Niu, Kai Cai, and Simon Parsons. *A Grey-Box Approach to Automated Mechanism Design*, pages 47–61. Springer, Berlin, Heidelberg, 2012.

[72] Jinzhong Niu, Kai Cai, Simon Parsons, Enrico Gerding, Peter McBurney, Thierry Moyaux, Steve Phelps, and David Shield. JCAT: A platform for the TAC market design competition. *Proceedings of the 7th international joint conference on autonomous agents and multiagent systems*, pages 1649–1650, 2008.

[73] Jinzhong Niu, Kai Cai, Simon Parsons, and Elizabeth Sklar. Some preliminary results on competition between markets for automated traders. *AAAI-07 Workshop on Trading Agent*, pages 19–26, Jul. 2007.

[74] James Norris. *Markov chains*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge, UK, 1998.

[75] Martin Nowak. *Evolutionary Dynamics*. Belknap Press, Cambridge, MA, 2006.

[76] Marco Pangallo, James Sanders, Tobias Galla, and J Doyne Farmer. A taxonomy of learning dynamics in 2× 2 games. *arXiv preprint arXiv:1701.09043*, 2017.

[77] Simon Parsons, Marek Marcinkiewicz, Jinzhong Niu, and Steve Phelps. Everything you wanted to know about double auctions, but were afraid to (bid or) ask. Technical report, University of Liverpool, 2006.

[78] William Press, Saul Teukolsky, William Vetterling, and Brian Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, New York, NY, USA, 3 edition, 2007.

[79] Kali Rath. A direct proof of the existence of pure strategy equilibria in games with a continuum of players. *Economic Theory*, 2(3):427–433, 1992.

[80] Hannes Risken. *The Fokker–Planck Equation*. Springer, Berlin, 1984.

[81] Edward Robinson, Peter McBurney, and Xin Yao. Co-learning segmentation in marketplaces. In *Lecture Notes in Computer Science*, volume 7113 LNAI, pages 1–20. Springer, Berlin, Heidelberg, 2012.

[82] Carlos Roca, José Cuesta, and Angel Sánchez. Time scales in evolutionary dynamics. *Physical Review Letters*, 97(15):158701, 2006.

[83] Carlos Roca, José Cuesta, and Angel Sánchez. Evolutionary game the-

ory: Temporal and spatial effects beyond replicator dynamics. *Physics of Life Reviews*, 6(4):208–249, 2009.

[84] Tim Rogers and Alan McKane. A unified framework for schelling's model of segregation. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(07):P07006, 2011.

[85] Egle Samanidou, Elmar Zschischang, Dietrich Stauffer, and Thomas Lux. Agent-based models of financial markets. *Reports on Progress in Physics*, 70(3):409, 2007.

[86] Yuzuru Sato, Eizo Akiyama, and James Crutchfield. Stability and diversity in collective adaptation. *Physica D*, 210:21–57, 2005.

[87] Yuzuru Sato, Eizo Akiyama, and Doyne Farmer. Chaos in learning a simple two-person game. *Proceedings of the National Academy of Sciences*, 99(7):4748–4751, 2002.

[88] Yuzuru Sato and James Crutchfield. Coupled replicator equations for the dynamics of learning in multiagent systems. *Physical Review E*, 67:015206, 2003.

[89] Thomas Schelling. Dynamic models of segregation. *The Journal of Mathematical Sociology*, 1(2), 1971.

[90] David Schmeidler. Equilibrium points of nonatomic games. *Journal of Statistical Physics*, 7(4):295–300, 1973.

[91] Bing Shi, Enrico Gerding, Perukrishnen Vytelingum, and Nick Jennings. An equilibrium analysis of market selection strategies and fee strategies in competing double auction marketplaces. *Autonomous Agents and Multi-Agent Systems*, 26(2):245–287, 2013.

[92] Herbert Alexander Simon. A behavioral model of rational choice. *Quarterly Journal of Economics*, 69(1):99–118, 1955.

[93] Brian Skyrms. Chaos in game dynamics. *Journal of Logic, Language and Information*, 1(2):111–13, 1992.

[94] Richard Sutton and Andrew Barto. *Introduction to Reinforcement Learning*. MIT Press, Cambridge, MA, 1st edition, 1998.

[95] Igor Swiecicki, Thierry Gobron, and Denis Ullmo. "phase diagram" of a mean field game. *Physica A*, 442:467–485, Jan. 2015.

[96] Igor Swiecicki, Thierry Gobron, and Denis Ullmo. Schrödinger approach to mean field games. *Physical Review Letters*, 116(12):128701, Mar. 2016.

[97] Peter Taylor and Leo Jonker. Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, 40(1):145 – 156, 1978.

[98] Bence Tóth, Enrico Scalas, Jürgen Huber, and Michael Kirchler. The value of information in a multi-agent market model - the luck of the uninformed. *The European Physical Journal B*, 55(1):115–120, 2007.

[99] Arne Traulsen, Jens Christian Claussen, and Christoph Hauert. Coevolutionary dynamics: from finite to infinite populations. *Physical Review Letters*, 95(23):238701, 2005.

[100] Arne Traulsen and Christoph Hauert. Stochastic evolutionary game dynamics. *Reviews of nonlinear dynamics and complexity*, 2:25–61, 2009.

[101] Karl Tuyls, Dries Heytens, Ann Nowe, and Bernard Manderick. *Extended replicator dynamics as a key to reinforcement learning in multi-agent systems*, pages 421–431. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[102] Karl Tuyls, Pieter Jan'T Hoen, and Bram Vanschoenwinkel. An evolutionary dynamical analysis of multi-agent learning in iterated games. *Autonomous Agent Multi Agent Systems*, 12(1):115–153, 2006.

[103] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974.

[104] Nicolaas Van Kampen. *Stochastic processes in physics and chemistry*. North Holland, Amsterdam, The Netherlands, 3 edition, 1997.

[105] Fernando Vega-Redondo. *Economics and the Theory of Games*. Cambridge University Press, Cambridge, UK, 2003.

[106] Johannes Wohlmuth and Jørgen Vitting Andersen. Modelling financial markets with agents competing on different time scales and with different amount of information. *Physica A: Statistical Mechanics and its Applications*, 363(2):459 – 468, 2006.

[107] Michael Wooldridge. Introduction to Multiagent Systems. *Information Retrieval*, 30:152–183, 2002.